

OM of: US-09-303-518d-571 to: SwissProt_40:*

Date: Jun 30, 2002 8:44 AM

About: Results were produced by the Gencore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL=frame+np.model -DEV=xih
-O=/cgn2/USPROT.spool/US09303518/runat_28062002_142714_4362/app.query.fasta_1.23501
-DB=SwissProt_40 -OPMT=fastan -SUFFIX=isp -GAPOP=12.000
-GAPX2=4.000 -MINMATCH=0.100 -LOOPEL=0.000 -LOOPEXT=0.000
-GAPOP=6.500 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blomscore2
-TRN=human40.cdi -LIST=100 -DOCALLIGN=200 -THR_SCORE=pt
-THR_MAX=100 -THR_MIN=0 -ALIGN=45 -MODE=LOCAL -OUTFMT=pts
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USBR=US09303518 @CGN1.1.440 -NCPU=6 -ICPU=3 -LONGLOG
-DEV_TIMEOUT=120 -NARN_TIMEOUT=30 -NO_XLUPY -WAIT -THREADS=1

Search information block:

Query: US-09-303-518d-571
Query length: 894
Database: SwissProt_40:*\br/>Database sequences: 105224
Search time (sec): 217.960000

score_list:

Sequence	Strd Orig	ZScore	EScore	Len	Documentation
SwissProt_40:HTB1.ECOLI +	196.50	272.97	5.4e-08	306	P24187 <i>Escherichia coli</i> , ddd
SwissProt_40:IDG.ECOLI +	168.00	232.26	9.9e-06	318	P45672 <i>Escherichia coli</i> , ddd
SwissProt_40:MSB.HAEN +	139.00	191.37	0.0019	311	P45672 <i>Escherichia coli</i> , ddd
SwissProt_40:SGS.DROE +	126.50	173.20	0.0182	328	P13730 <i>Haemophilus influenza</i>
SwissProt_40:YGR3.HALO +	125.00	171.21	0.0238	323	P24205 <i>Escherichia coli</i> , ddd
SwissProt_40:MSB.HAEN +	122.50	165.04	0.0369	337	P21361 <i>haloflex</i> sp. (Strat)
SwissProt_40:NEP1.HUMAN +	119.00	167.75	0.0519	231	P06659 <i>shigella flexneri</i> , 11
SwissProt_40:HRK3.HUMAN +	117.50	152.02	0.0818	1104	P10298 <i>homo sapiens</i> (human)
SwissProt_40:DBA.RAT +	110.50	149.70	0.1017	688	P10074 <i>homo sapiens</i> (human)
SwissProt_40:VE2.HPV1 +	110.50	146.99	0.3365	361	P06764 <i>rattus norvegicus</i> (rat)
SwissProt_40:MEP1.LOPE +	108.50	148.25	0.4752	493	P26766 <i>rattus norvegicus</i> (rat)
SwissProt_40:YFRA.BORP +	108.50	147.92	0.4773	308	P04024 <i>loium perenne</i> (pete)
SwissProt_40:VE2.HPV4 +	108.00	143.22	0.5512	506	P23445 <i>bordelella pertussis</i>
SwissProt_40:GLT5.WHEAT +	107.00	137.41	0.7009	839	P10388 <i>tritium aestivum</i> (wt)
SwissProt_40:ICP0.HSV1 +	105.50	146.89	0.7886	221	P062093 <i>mus musculus</i> (mouse)
SwissProt_40:CCA.RABIT +	105.50	126.05	1.04	2424	P27884 <i>oryctolagus cunicul</i>
SwissProt_40:SON.HUMAN +	105.00	126.05	1.04	2426	P18583 <i>homo sapiens</i> (human)
SwissProt_40:YFRA.BORP +	105.00	142.41	0.9081	253	P18824 <i>aplysia californica</i>
SwissProt_40:CFE.CHLV +	105.00	142.41	0.9081	341	P56316 <i>rattus norvegicus</i> (rat)
SwissProt_40:VE2.HPV1 +	105.00	134.72	1.01	825	P06303 <i>rattus norvegicus</i> (rat)
SwissProt_40:REP1.MOUSE +	105.00	131.98	1.04	1130	P97347 <i>mus musculus</i> (mouse)
SwissProt_40:YF44.STRE +	104.50	138.13	1.01	395	P20188 <i>streptomyces fradiae</i>
SwissProt_40:SON.HUMAN +	103.50	123.29	1.50	514	P06921 <i>human papillomavirus</i>
SwissProt_40:SR22.CHICK +	102.50	132.64	1.36	2404	P09647 <i>mus musculus</i> (mouse)
SwissProt_40:HR22.HUMAN +	102.50	132.64	1.36	221	P03352 <i>gallus gallus</i> (chick)
SwissProt_40:VE2.HPV3 +	102.00	134.67	1.55	673	P51116 <i>homo sapiens</i> (human)
SwissProt_40:VE2.HPV3 +	102.00	132.41	1.69	509	P50809 <i>human papillomavirus</i>
SwissProt_40:SGS4.DROME +	101.50	138.65	1.66	660	P03181 <i>epstein-barr virus</i> (str)
SwissProt_40:MANA.RHOM +	101.50	137.91	1.95	297	P000725 <i>strepococcus</i>
SwissProt_40:VE2.HPV5 +	101.00	133.17	1.97	1021	P26445 <i>rhodotermus marinus</i>
SwissProt_40:ITLA.MOUSE +	100.50	136.16	2.05	514	P26445 <i>rhodotermus marinus</i>
SwissProt_40:ITLA.MOUSE +	100.50	129.98	2.23	336	P11088 <i>mus musculus</i> (mouse)
SwissProt_40:ITLA.MOUSE +	100.00	133.60	2.30	684	P10600 <i>aeromonas hydrophila</i>
SwissProt_40:CCAB.RABIT +	100.00	118.57	2.82	416	P20930 <i>homo sapiens</i> (human)
SwissProt_40:CCAB.RABIT +	99.00	129.11	2.88	2339	P05152 <i>oryctolagus cunicul</i>
SwissProt_40:CCAB.HUMAN +	99.00	117.97	3.38	592	P09502 <i>homo sapiens</i> (human)
SwissProt_40:CCAB.HUMAN +	98.50	116.16	2.81	2339	P00975 <i>homo sapiens</i> (human)
SwissProt_40:CCAB.HUMAN +	98.50	116.16	2.81	221	P001130 <i>homo sapiens</i> (human)

SwissProt_40:VE2.HPV3 +	97.00	128.59	4.02	454	P08093 <i>human papillomavirus</i>
SwissProt_40:ITB3.HSVH +	97.00	127.54	4.07	512	P28276 <i>herpes simplex vir</i>
SwissProt_40:YB4.SCHO +	96.50	131.57	4.19	297	P10357 <i>schistosoma</i>
SwissProt_40:CCRI.NEUC +	96.00	127.64	4.79	430	P05958 <i>neurospora crassa</i>
SwissProt_40:NONA.DROME +	95.50	122.69	5.55	700	P04047 <i>drosophila melanog</i>
SwissProt_40:VE2.HPV17 +	95.00	125.79	5.77	452	P36785 <i>human papillomavirus</i>
SwissProt_40:SRB.HUMAN +	94.50	124.49	6.37	484	P05519 <i>homo sapiens</i> (huma
SwissProt_40:ZKRI.CHICK +	94.50	122.75	6.52	591	P37767 <i>gallus gallus</i> (chl
SwissProt_40:UHB1.USTWA +	94.50	118.62	6.89	950	P40344 <i>usliago maydis</i> (s
SwissProt_40:DBA.HUMAN +	94.00	126.07	6.77	372	P16983 <i>homo sapiens</i> (huma
SwissProt_40:GLT4.HUMAN +	93.50	127.90	7.17	838	P08489 <i>tritium aestivum</i>
SwissProt_40:YF4.RHIS +	93.00	119.82	8.66	278	P17968 <i>rhizobium sp.</i> inser
SwissProt_40:SPRA.MOUSE +	93.00	108.39	10.08	2411	P16383 <i>homo sapiens</i> (human)
SwissProt_40:GTO.HUMAN +	93.00	108.39	10.08	2411	P16383 <i>homo sapiens</i> (human)
SwissProt_40:GCF.HUMAN +	93.00	113.15	9.46	1396	P12080 <i>drosophila melanog</i>
SwissProt_40:THA2.DROME +	93.00	113.15	9.46	1396	P12080 <i>drosophila melanog</i>
SwissProt_40:DBA.DROME +	93.00	108.39	10.08	2411	P16383 <i>homo sapiens</i> (human)
SwissProt_40:CCAA.HUMAN +	92.50	129.43	8.26	198	P40325 <i>saccharomyces cere</i>
SwissProt_40:YGS4.YEAST +	92.50	122.56	9.05	436	P08170 <i>homo sapiens</i> (huma
SwissProt_40:SR4.HUMAN +	92.50	121.48	9.18	436	P08170 <i>homo sapiens</i> (huma
SwissProt_40:YK82.YEAST +	92.50	112.82	10.15	1169	P36170 <i>saccharomyces cer</i>
SwissProt_40:WMP4.HUMAN +	92.50	111.93	11.32	1336	P15399 <i>homo sapiens</i> (hum
SwissProt_40:SU22.DROME +	92.00	125.54	10.24	1365	P25172 <i>drosophila melano</i>
SwissProt_40:SGS3.DROYA +	91.50	120.06	11.01	263	P13728 <i>drosophila yakuba</i>
SwissProt_40:VE2.HPV12 +	91.50	119.82	11.05	494	P36782 <i>human papillomavirus</i>
SwissProt_40:GATB.BOMO +	91.50	117.79	11.35	508	P32167 <i>bombyx mori</i> (silik
SwissProt_40:EBN1.EBV +	91.00	125.63	11.09	641	P03211 <i>epstein-barr virus</i> (
SwissProt_40:RL32.HAMA +	91.00	123.78	11.37	240	P12736 <i>halocaula marino</i>
SwissProt_40:YK22.TREH +	91.00	119.28	12.07	297	P01200 <i>leptocaula edodes</i> (
SwissProt_40:VE2.HPV8 +	91.00	109.96	13.72	498	P06422 <i>treponema pallidum</i>
SwissProt_40:ATY4.HUMAN +	91.00	107.96	14.04	1500	P04656 <i>homo sapiens</i> (hum
SwissProt_40:Y236.HUMAN +	90.00	122.02	13.70	1829	P015015 <i>homo sapiens</i> (hum
SwissProt_40:U2R1.HUMAN +	90.00	118.20	14.41	309	P17143 <i>human cytomegalovir</i>
SwissProt_40:VE2.HPV4 +	89.50	118.04	14.45	479	P15695 <i>homo sapiens</i> (huma
SwissProt_40:VE2.HPV21 +	89.50	117.07	15.87	503	P26795 <i>human papillomavirus</i>
SwissProt_40:CPM4.HUMAN +	89.50	113.69	16.61	742	P16070 <i>h cd4 antigen pre</i>
SwissProt_40:POLN.DROME +	89.50	107.37	18.06	1533	P12822 <i>drosophila melanog</i>
SwissProt_40:POLN.DROME +	89.50	103.19	19.10	2479	P12822 <i>drosophila melanog</i>
SwissProt_40:Y2A1.ADEO2 +	89.50	96.46	20.89	5176	P08799 <i>mus musculus</i> (mouse)
SwissProt_40:Y2A1.ADEO2 +	89.00	123.76	15.75	215	P08799 <i>mus musculus</i> (mouse)
SwissProt_40:ICP0.HSV1 +	89.00	118.85	16.82	318	P03691 <i>human adenovirus</i> t
SwissProt_40:YF44.STRE +	88.50	124.64	16.89	775	P08393 <i>herpes simplex vir</i>
SwissProt_40:YF44.STRE +	88.50	110.77	20.32	882	P15265 <i>mus musculus</i> (mouse)
SwissProt_40:YF44.STRE +	88.50	104.89	21.97	1733	P01741 <i>borrelia burgdorfer</i>
SwissProt_40:YF44.STRE +	88.00	121.98	18.98	224	P08287 <i>gallus gallus</i> (chi
SwissProt_40:YF44.STRE +	88.00	118.46	19.80	336	P12574 <i>xenopus laevis</i> (af
SwissProt_40:YF44.STRE +	88.00	112.58	21.51	660	P08489 <i>tritium aestivum</i>
SwissProt_40:YF44.STRE +	88.00	112.02	21.68	704	P17777 <i>rattus norvegicus</i> (r
SwissProt_40:YF44.STRE +	88.00	109.29	22.48	963	P09457 <i>ceenorhabdilis ele</i>
SwissProt_40:YF44.STRE +	88.00	104.92	23.83	1593	P15837 <i>homo sapiens</i> (hum
SwissProt_40:YF44.STRE +	87.50	113.33	23.11	558	P00257 <i>homo sapiens</i> (huma
SwissProt_40:YF44.STRE +	87.50	109.04	24.47	914	P08365 <i>escherichia coli</i> o
SwissProt_40:YF44.STRE +	87.50	108.33	24.70	992	P08563 <i>rubella virus</i> (str

seq_name: SwissProt_40:HTB1.ECOLI
seq_documentation_block:
ID: HTB1.ECOLI STANDARD: PRT: 306 AA.
AC: P24187
DT: 01-MAR-1992 (Rel. 21, Created)
DT: 01-MAR-1992 (Rel. 21, Last sequence update)
DT: 16-OCT-2001 (Rel. 40, Last annotation update)
DE: Lipid A biosynthesis, lauroyl acyltransferase (EC 2.3.1.-) (Heat shock protein B).
GN: HTB1 OR WAAW OR B1054 OR Z1690 OR ECG1432.
OS: *Escherichia coli*, and

[illegible]

seq_name: SwissProt_40:DDG_ECOLI

```

seq_documentation_block:
ID      DDG_ECOLI      STANDARD:      PRT:      306 AA.
AC      P76522; P76949;
DT      01-NOV-1997 (Rel. 35, Created)
DT      01-NOV-1997 (Rel. 35, Last sequence update)
DT      16-OCT-2001 (Rel. 40, Last annotation update)
DE      DDG protein.
GN      DDG OR B2378.
OS      Escherichia coli.
OC      Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC      Escherichia.
OX      NCBI_TaxID=562;
       [1]
       SEQUENCE FROM N.A.
RP      STRAIN=K12 / W3110;
RA      Sreekumar K.R., Schechter M.;
RL      Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
RN      [2]
       SEQUENCE FROM N.A.
RC      STRAIN=K12 / MG155;
RX      MEDLINE=97426617; PubMed=9278503;
RA      Blatner F.R., Plunkett G. II, Bloch C.A., Perna N.T., Burland V.,
RA      Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA      Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA      Mau B., Shao Y.;
       "The complete genome sequence of Escherichia coli K-12.";
       Science 277:1453-1474.(1997).
       [3]

```

```

RP SEQUENCE FROM N.A.
RX STRAIN-K12:
RX MEDLINE=97349980; PubMed=9205837;
RA Yamamoto Y., Alba H., Baba T., Hayashi K., Inada T., Isono K.,
RA Itoh T., Kimura S., Kitagawa M., Makino K., Miki T., Mitsunashi N.,
RA Mizuduchi K., Mori H., Nakade S., Nakamura Y., Nishimoto H.,
RA Oshima T., Oyama S., Saito N., Saito G., Sato Y., Sivasubraman S.,
RA Tagami H., Takahashi H., Takeda J., Takemoto K., Uehara K., Wada C.,
RA Yamagata S., Horinouchi T.;
RT "Construction of a contiguous 874-kb sequence of the Escherichia coli
RT -K12 genome corresponding to 50-0:68.8 min on the linkage map and
RT analysis of its sequence features.";
RL DNA Res. 4:91-113(1997).
CC -I- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
CC (potential).
CC -I- SIMILARITY: BELONGS TO THE HTRB/MSB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; U49787; AAB66658.1; -
DR EMBL; AE000326; AAC75437.1; ALT INIT.
DR EMBL; D90869; BAA16248.1; -.
DR Ecogene; EGI2901; ddd.
KW Transmembrane; Inner membrane; Transferase; Acyltransferase;
KW Complete proteome.
FT TRANSMEM 20 40 POTENTIAL.
FT TRANSMEM 125 145 POTENTIAL.
SQ SEQUENCE 306 AA; 35493 MW; 58327F3D96C9E7B6 CRC64;

```

[illegible]

Align seg 1/1 to: DG_ECOLI from: 1 to: 306

```

12 CAATAGCCCTCCTGCTGCTGCTTCTGTCGACACGGCTGGAAACC 121
13 GlnLeuProTyrProValLeuCysPheLeuGlyThrArgIleGlyAlaI 47
14 122 GGCCTCGGACATCTGGCGTTTACCTTTTAAAGAAGACCGCGCGCATC 171
15      |||||      |||||      |||      :|||
16 47 TAlaArg.....ProPheLeuLysArgArgGlnSerIleA 59
17 172 GTGCG.....CAATATGCGGACGGCGGTTT 197
18      |||      |||||      |||||      |||
19 59 TaaArgLysAsnLeuGlnLeuCysPheProGlnHisSerIla..... 72
20 198 GAACCCCGACACGACGACGGTCAAAAGCGTTTTCGGGAACCGCAAAAT 247
21      |||      |||      :|||      |||      :|||
22 73 GlnGluArgGlu..LysMetIleAlaGlnAsnPheArgSerLeuGlyMetA 89
23 248 GCGGTTTGGAACTTGCCCCGCGCTTTTTCAAAAACCGAAGACATCGAA 297
24      :|||      :|||      :|||      :|||
25 89 TAlauValGlnThrGlyMetAlaTrpPheThrProAspSerArgValArg 105
26 298 ACAATGTTCAACCGGTACACGGCTGGGAACACGTGACACGAGCTTTTGA 347
27      |||      |||||      :|||      :|||
28 106 LysTrpPheasp...ValGlnGlyLeuAspAsnLeuLysArgAlaGlnMe 121
29 348 CAAGGCGCAAGGGCTGCTGTCATTCACGCGGACATCGGCACTACGATT 397
30      :|||      :|||      :|||      :|||
31 121 tGlnAsnArgGlyValMetValValGlyAlaHisPheMetSerLeuGlnL 138

```

```

398 TGGGGGAGCGCTACATGACGACGACGCTTCGTCACATGACGCGCATG 447
|||||
138 euglyglylvalmetgylleucysglnpro.....MetMetAlaThr 152
|||||
448 TACAAAGCGCGCGCAAAATGACGATAGCAAAATGATGACGAGGCGGAC 497
|||||
153 tyrtatgprphisasmnslndleumetglutrpvalglnthrargglylar 169
|||||
498 GGTGCGCGGCAAAAGCAAAACCGCGCCACGCGCATACAAAGGGGTCAAC 547
|||||
169 gmetarg.....SerAsnlysalmetlledlyargasmnslndleuargg 184
|||||
548 AATATCATGAGCGCGCGCGCGCGGAGCAACCATCATCTGCGCGCAC 597
|||||
184 lytlevalglalaleuylslysglylualavaltrpheapalarpasp 200
|||||
598 CACGTCCTCTCCGAGAGGCGCGCGCGGTGCGGCGGATTTTTCGG 647
|||||
201 Gln.....AsptryglvalrlyglyserSerPheAlaProPhePheAl 215
|||||
648 C.....AAACGTCATACACCATGACACTGCGGCAAAATTTGGACACATCA 694
|||||
215 avalgluasnvalalathrThrAsnlylthryrvalleuserargleus 232
|||||
695 AAGGCGGTGAAACCGTGTTCCTGCTGCGAACCGCTGCCGACGAGACA 744
|||||
232 erglyalalaleuethrvalthrmetvalrlysalaspyrser 248
|||||
745 GCGTCGCTGTTCGACATCCGCGCGCGCGCAAGGGAATTAACGCGC.... 789
|||||
249 glytylthrleuphelethrpro.....Glnmetgluglytyrtr 262
|||||
790 .....ACAAAGCCGACGATGCGCGCGGTTCACCGCAATACGC 829
|||||
262 othrAspgluasnlnla.....AlaAlaTyrmetslnlysllelelg 277
|||||
830 AATATGTGATACGCGCTTCGCGACGATGCTGTTATGATGACACCGC 879
|||||
277 lulysglulilemetargalalProgluInlyrleuprlehlslargarg 293
|||||
880 TATAAAGC 888
|||||
294 Pheylstr 296
|||||
seq_name: SwissProt_40:MSBB_HAEIN
seq_documentation_block:
ID MSBB_HAEIN STANDARD; PRT; 318 AA.
AC P44567;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Lipid A biosynthesis (KDO)2-(lauroyl)-lipid IVA acyltransferase
DE (EC 2.3.1.-)
GN MSBB OR H10199.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OC NCBTaxid=727;
OX [1]
RN
RP SEQUENCE FROM N.A.
RC STRAIN=RD / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kertevage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McInerney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shiley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Uetlerback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;

```

```

RT "Whole-genome random sequencing and assembly of Haemophilus
RT Influenzae Rd."
RL Science 269:496-512(1995).
CC -1- FUNCTION: TRANSFERS MYRISTATE OR LAURATE, ACTIVATED ON ACP, TO
CC (KDO)2-(LAUROYL)-LIPID IVA (BY SIMILARITY).
CC -1- PATHWAY: LIPIDPOLYSACCHARIDE CORE BIOSYNTHESIS (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
CC (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE HTRB/MSBB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-stb.ch).
CC -----
CC EMBL; U32705; AAC21868.1; -.
CC TIGR; H10199; -.
CC Lipopolysaccharide biosynthesis; Transferase; Acyltransferase;
CC Transmembrane; Inner membrane; Complete proteome.
CC TRANSMEM 27 47 POTENTIAL.
CC TRANSMEM 91 111 POTENTIAL.
CC TRANSMEM 138 158 POTENTIAL.
CC SEQUENCE 318 AA; 36882 MW; DE59952D78719445 CRC64;

alignment_scores:
Quality: 168.00 Length: 265
Ratio: 1.105 Gaps: 10
Percent Similarity: 57.358 Percent Identity: 24.151

alignment_block:
US-09-303-518D-571 x MSBB_HAEIN ..
Align seg 1/1 to: MSBB_HAEIN from: 1 to: 318

112 CTGGGAACCGCGCTCGACATCTGCGGCTTTTACCTTTAAGAGACCG 161
|||||
57 Lendlylletprlleglyhsllysala.....Lyslyglaln 69
|||||
162 CGCGCGCATGCTGCGCAATATGCGGCGCG.....GTTTGA 199
|||||
69 gthrArgAlaGlnThrAsnleuglnlyrcysrPheonlstrPrhngln 86
|||||
200 ACCCGGACGCGCGCGCTGCAAGCGCTTTTGGCGAAGCGCAAAATGC 249
|||||
86 lnglnArggluInvalIleAspIysmetPheAlaValAlaIleGlnval 102
|||||
250 GCTTGGAACTTGCGCGCGCGCTTTTCAAAAACCGGAAACATGCAAC 299
|||||
103 MetPhegllyleuglylualleAlaIleArgserlyshslsleuInly 119
|||||
300 AATGTTCAAGCGGTACAGCGCTGGGAACAGCGGACGCGCTTGACA 349
|||||
119 sargsergluphele.....GlyleuglnslsleleuglnAlalyala 135
|||||
350 AGGGGGAAGGGGTGCTGTCATGACGCGCGACATGCGCGACGATTTG 399
|||||
135 lueglylvasnlelleleumetvalProhissglytrAlaIleAspAla 151
|||||
400 GCGGAGCGCTACATC...AGCGACAGCTTCGCTTCACCTGACCGGCAT 446
|||||
152 serglyllelleleuInlThrGlnlymetPro.....MetThrSerMe 166
|||||
447 GPACAAAGCGCGCGCAAAATGACGATGACAAATGATGACGAGGCGGCA 496
|||||
166 tyrtaspronlslargasnProleuValAspTrpleutrPrhrllehra 183
|||||
497 GGTGCGCGGCAAAAGCAAAACCGCGCGCGGATACAAAGGGGTCAAA 546
|||||
183 rglInArgPhegllyllysmethslAlaArgGln...Asnlyllelys 198
|||||

```


716 TCTGCTGGACGCGCTGCCGAGACAGCTTGTGTGACATCCGC 765
 247 ..Thrllysargalathrllysargalathr...Thrlsargal 261
 766 CCGCTCCAGGAGGATTTGAACG.....CCACAAAGCCGACGAGC 806
 261 aProthllysargalathrllysargalathrlalathrlSerl 278
 807 CGCGCTGT 814
 278 ysrProcys 280

seq_name: SwissProt_40:MSBB_ECOLI

seq_documentation_block:

AC MSBB_ECOLI STANDARD; PRT; 323 AA.

DI 01-MAR-1992 (Rel. 21, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DE Lipid A biosynthesis (KDO)2-(lauroyl)-lipid IVA acyltransferase
 GN (EC 2.3.1.-)
 OS MSBB OR B1855.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 OX NCBI_Taxid=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12 / W3110;
 RX MEDLINE-92121107; PubMed-1732206;
 RA Karow M., Georgopoulos C.;
 RT "Isolation and characterization of the Escherichia coli msb gene, a
 RT multiplicity suppressor of null mutations in the high-temperature
 RT requirement gene htrB.";
 RL J. Bacteriol. 174:702-710(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12;
 RX MEDLINE-93015688; PubMed-1356966;
 RA Engel H., Slink A.J., van Wijngaarden L., Keck W.;
 RT "Murein-metabolizing enzymes from Escherichia coli: existence of a
 RL J. Bacteriol. 174:6394-6403(1992).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12 / M61655;
 RX MEDLINE-97426617; PubMed-9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of Escherichia coli K-12.";
 RL Science 277:1453-1474(1997).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12;
 RX MEDLINE-97251358; PubMed-9097040;
 RA Itoh T., Aiba H., Baba T., Fujita K., Hayashi K., Inada T.,
 RA Isono K., Kasai H., Kimura S., Kitakawa M., Kitagawa M.,
 RA Makino K., Miki T., Mizobuchi K., Mori H., Motomura K.,
 RA Nakade S., Nakamura Y., Nishimoto H., Nishio Y., Oshima H.,
 RA Saito N., Sampei G., Seki Y., Sivasubram S., Tagami H.,
 RA Takeda J., Takemoto K., Wada C., Yamamoto Y., Horikuchi T.;
 RT "A 460-kb DNA sequence of the Escherichia coli K-12 genome
 RT corresponding to the 40.1-50.0 min region on the linkage map.";
 RL DNA Res. 3:379-392(1996).
 RN [5]
 RP FUNCTION AND CHARACTERIZATION.
 RX MEDLINE-97256743; PubMed-9099672;
 RA Clementz T., Zhou Z., Raetz C.R.H.;
 RT "Function of the Escherichia coli msbB gene, a multiplicity suppressor
 RT of htrB knockouts, in the acylation of lipid A. Acylation by msbB

RT follows laurate incorporation by HtrB.";
 RL J. Biol. Chem. 272:10353-10360(1997).
 CC -1- FUNCTION: TRANSFERS MYRISTATE OR LAURATE, ACTIVATED ON ACP, TO THE
 CC LIPID IVA MOIETY OF (KDO)2-(LAUROYL)-LIPID IVA. DECANOYL,
 CC PALMITOYL, PALMITOLEOYL, AND (R)-3-HYDROXYMYRISTOYL-ACP ARE POOR
 CC ACYL DONORS. FUNCTIONS OPTIMALLY AFTER LAURATE INCORPORATION BY
 CC HTRB HAS TAKEN PLACE. ACYLATES (KDO)2-LIPID IVA. DISPLAYS A PREFERENCE FOR
 CC 100 TIMES FASTER THAN (KDO)2-LIPID IVA. ABOUT
 CC MYRISTOYL-ACP OVER LAUROYL-ACP.
 CC -1- PATHWAY: LIPIDOLYSACCHARIDE CORE BIOSYNTHESIS.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane.
 CC -1- SIMILARITY: BELONGS TO THE HTRB/MSBB FAMILY.
 CC -1- CAUTION: WAS ORIGINALLY (REF.2) THOUGHT TO BE THE MEMBRANE-BOUND
 CC LYTIC MUREIN TRANSGLYCOSYLASE (MLT).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>
 CC or send an email to license@isb.slb.ch).
 CC -----
 DR EMBL: M77039; AAA24181.1;
 DR EMBL: M87660; AAA96706.1;
 DR EMBL: AE000279; AAC74925.1;
 DR EMBL: D90828; BAA15663.1;
 DR PIR: A42608; A42608.
 DR Ecogene; EG10614; msbB.
 KW Lipidolysaccharide biosynthesis; Transferase; Acyltransferase;
 KW Transmembrane; Inner membrane; Complete proteome.
 FT TRANSMEM 23 43
 FT TRANSMEM 85 105 POTENTIAL.
 FT TRANSMEM 133 153 POTENTIAL.
 SO SEQUENCE 323 AA; 37410 MW; 94DAP38A757D20CD CRC64;

alignment_scores:

Quality: 125.00 Length: 274
 Percent Similarity: 0.899 Gaps: 11
 Percent Identity: 50.730 Percent Identity: 24.088

alignment_block:

US-09-303-518D-571 x MSBB_ECOLI ..

Align seg 1/1 to: MSBB_ECOLI from: 1 to: 323

121 CGGCTCGGACATCTGCGCTTTACCTTTAAAGAGACCGCGCGCAT 170
 |||||
 50 ArgLeuGlyArgPheAlaGlyArgLeuGlyLysSerSerArgArgArg 66
 171 CCGCGCAATATG.....CGGC 187
 66 aLeuLeuLeuSerLeuGlySerPheProGluArgSerGluAlaGluArg 83
 188 AGCGGCGTTGAACCCGACGACGACGATCAAGCGCTTTTCGGA 237
 83 IuAla.....IleValAspIlePheAlaThr 92
 238 ACAGCAATGCGGTTTGAACCTGCGCGCGGTTTTCAAAAACCGGA 287
 :|||
 93 AlaProGluAlaMetAlaMetMetAlaGluAlaIleArgGlyProG 109
 288 AGACATCGAACAATGTTCAAGCGCTGACGCGCGGACAGTCGAGC 337
 |||||
 109 uGylIleGlnProArgValAsp...TrpGlnGlyLeuGluIleLeuG 125
 338 AGGCTTTGACAAAGGCGGAGGCGCTGTCATCAGCGGACATCGCG 387
 :|||
 125 IuMetArgArgAsnGlnIuYsValIlePheLeuValProHisGlyTrp 141
 388 AGCTACGATTTGGCGGACGCTACATCAGCAGCAGCTTCGTTCACT 437
 :|||
 :|||

142 AlayaValaIleProAlaMetLeuMetaIaSerGln...GlyInLysLe 157
 438 GACCGCCATGTTACAAAGCCGCAAAATCAAGGATACAAATATCATATC 487
 157 ValAlaMetPheHisSngInglYasnProValPheAspTyrValTyrPA 174
 488 AGCGGGGAGGAGTGGCGGCAAGCAAAACCGCGCCACCGCATACAA 537
 174 snThrValArGArGArPheGlyGlyArGLeuHisAlaArGasn...Asp 189
 538 GGGGTCAAACAATCATCAAGCCCTGCGCGCGCGGCAAGCAACATCAT 587
 190 GlyLeuYsrProPheIleGlnSerValArGInglYrTyrTrpGlyTyr 206
 588 CTGCCC.....GACCACTCCCTTCGCCGAGAGAGCGCGCGCTGT 631
 206 rLeuProAspGlnAspHisGlyProGlnHisSerGlu.....P 219
 632 GGGGGATTTTTCGCAAACTGTCATACACATGACACTGGCGGCAAA 681
 219 heValAspPhePheAlaThrTyrLysAlaThrLeuProAlaIleGlyArG 235
 682 TTGGCAGACGTC.....AAGCGTGGAACCCCTGTTTCTGCTGCGA 725
 236 LeuMetLysValLysArGAlaArGValAlaProLeuPhe..... 248
 726 AGCGCTGCCGCGAGCAAA.....GGCTTCGTGTCACATCCGCGCGG 769
 249 rProIleYrAspGlyLysThrHisArGLeuThrIleGlnValArProP 265
 770 TCCAAAGGGATTTGAACGCAACAAAGCCGCAAT...GCCGCGCTGTTC 816
 265 roMetAspAspLeuLeuGlnAlaAspAspHisThrIleAlaArGArMet 281
 817 AACCGCAATACCGAATATTGATACGCGCTTTCCGACGCAATTCGT 866
 282 AsnGlnGluValGlnIlePheValGlyProArGProGlnTyrTrpTr 288
 867 TATGTACAAACCGCTATTAACG 888
 298 rIleLeuLysLeuLeuLysThr 305
 seq_name: SwissProt_40:YGY3_HAUSO
 seq_documentation_block:
 ID YGY3_HAUSO STANDARD: PRT: 437 AA.
 AC P21561;
 DT 01-MAY-1991 (Rel. 18, Created)
 DT 01-MAY-1991 (Rel. 18, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical 50.6 kDa protein in the 5' region of GYRA and GYRB (ORF
 3).
 OS Haloferax sp. (strain Aa 2.2).
 OC Archaea; Euryarchaeota; Halobacteriales; Halobacteriaceae; Haloferax.
 OX NCBI_TaxID=2254;
 RN [1]
 RP MEDLINE=91100352; PubMed=1846146;
 RA Holmes M.L., Dyal-Smith M.L.;
 RT "Mutations in DNA gyrase result in novobiocin resistance in
 J. Bacteriol. 173:642-648(1991)."
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See http://www.isb-sib.ch/announce/
 or send an email to license@isb.sib.ch).
 CC EMBL: M38373; -; NOT_ANNOTATED_CDS.
 DR PIR; C39135; C39135.

KW Hypothetical protein.
 SQ SEQUENCE 437 AA; 50626 MW; B5B99A2AF3892BEF CRC64;
 alignment_scores:
 Quality: 122.50 Length: 306
 Ratio: 0.996 Gaps: 12
 Percent Similarity: 40.196 Percent Identity: 23.529
 alignment_block:
 US-09-303-518D-571 x YGY3_HAUSO ..
 Align seg 1/1 to: YGY3_HAUSO from: 1 to: 437
 27 TCCGCGCTTCGGAACCGCCATGCATCTCTGTGACCGCCCTGCTCAAT 76
 130 SerArGLeuArGInGlnHisGlnHisProArGArGLeu..... 142
 77 GCGTCTCCCTGCTGCTGCTCTCTGCTGCTGCAACGCGTGC 126
 143HisAlaSerAspArGValG 149
 127 GCACATCTGGCGTTTACCTTTAAAGAGACCGCGCGCATGCTGCC 176
 149 In...AspGlyAlaHisProArGArGInArGLeuArGLeu..... 161
 177 CAATATGCGGAGCGCGGCTTGAACCCCGACGACGACGTCMAAGCGG 226
 162GlnProArGHisAla..... 166
 227 TTTTTCGGAACCGCAAAATGCGGTTTGAACCTTGGCCCGCGCTTTTTC 276
 167GlyArGProArGArGArGlnProProArGArGLeu 179
 277 AAAAACCAGAGACATCAACATATGTTCAAGCGGTACAGCGCTGGA 326
 179 rSerArGLeuThrHisArGArGHisLeuArGInAlaProArGProAla 195
 327 ACACGTCGACGACGCTTGGACAAAGCGGACGCGCTGCTGTCATCAGC 376
 196 ValArGLeuProAspGlnAspGlnAlaArGLeu..... 206
 377 CGCACATGCGACGATGATTTGGCGAGCGCTACATCAGCCAGAGCTT 426
 207PheArGLeuProArGHisArGArGLeuArGHisProProThrAla. 221
 427 CCGTCCACCTGACCGCATGTACAAAGCCGCGCAAAATCAAGCGATAGA 476
 222ArgAspValLeuArGLeuProGlnHisGlyAsp 233
 477 CAATATCATGACGCGGCGGAGGTGCGCGCAAAAGCGCAACCGCGCCA 526
 234 GlyHisHisLeuGlnGlyArG..... 240
 527 CCGGATACAAAGGGGTCAACAAATCATCAAGCGCTGCGCGGCGAG 576
 241ArgGlyArGProArGProGlnGlyArGLeuAlaGlyArG 254
 577 GCACATCATCTCT..... 590
 254 LysAlaHisProProGlnValArGAlaArGLeuTyrLeuAlaAlaGlyLeu 270
 591GCCGACCGAGTCCCTTC..... 608
 271 AlaArGLeuProGlnProArGProLeuGlyValArGThrValHisAr 287
 609TCCGAGAGAGCGCGCGCGCTG 630
 287 gGlyGlyArGLeuArGLeuArGValGlnAlaGlnAlaGlyProArGProGln 304
 631 TGGCGGATTTTTCG.....CAAAACGTCATACCATGACACT 671
 304 aProGlyAspPheAlaProGlnGlyGlnAspSerGlnArGArGLeuThr 320


```

672 GCGGCGCAAAATTTGGCAGCATCAAGGCGTGAACACCTGTTTCTGCT 721
    ::::::::::::::::::::
321 ProProArgProHisSerArgLysArgArgAspThrLeuAlaHisHisAr 337
722 GCGAAGCGCTCCCGGACGAGCAAGCGTCTG..... 752
    ::::::::::::::::::::
337 GHisTrpArgArgArgArgValArgHisArgGluGlyAlaLeuP 354
753 ..GTTGCACATCCGCGCCGTCGAAGGGAATTGAAGCGCAACCGCA 800
    ::::::::::::::::::::
354 roAlaAlaHisProAspArgArgArgArgArgArgArgAlaHisPro 370
801 CGATGCGCGCGCTGCTTC 816
    ::::::::::::::::::::
371 .AspAlaAlaAlaLys 375
seq_name: SwissProt_40:MSBB_SHIFL

```

```

seq_documentation_block:
ID MSBB_SHIFL STANDARD: PRT: 231 AA.
AC 00659:
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DE 01-MAR-2002 (Rel. 41, Last annotation update)
DE Lipid A biosynthesis (KDO)2-(lauroyl)-lipid IVA acyltransferase
  (EC 2.3.1.-) (Fragment).
GN MSBB.
OS Shigella flexneri.
OC Plasmid 230 kb pMYSH6000.
CC Bacteria: Proteobacteria; gamma subdivision; Enterobacteriaceae;
  Shigella.
OX NCBI_TaxID=623;
  [1]
  RP SEQUENCE FROM N.A.
  RC STRAIN=2A;
  RX MEDLINE=97315241; Pubmed-9711415;
  RA Rededge L., Davis M.A., Youngren B., Austin S.J.;
  RT "Plasmid maintenance functions of the large virulence plasmid of
  Shigella flexneri."
  RL J. Bacteriol. 179:3670-3675(1997).
  CC -1- FUNCTION: TRANSFERS MYRISTATE OR LAURATE, ACTIVATED ON ACP, TO
  (KDO)2-(LAUROYL)-LIPID IVA (BY SIMILARITY).
  CC -1- PATHWAY: LIPOLYSACCHARIDE CORE BIOSYNTHESIS (BY SIMILARITY).
  CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
  (Potential).
  CC -1- SIMILARITY: BELONGS TO THE HTRB/MSBB FAMILY.
  CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
  CC the European Bioinformatics Institute. There are no restrictions on its
  CC use by non-profit institutions as long as its content is in no way
  CC modified and this statement is not removed. Usage by and for commercial
  CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
  CC or send an email to license@isb-sib.ch).
  DR EMBL: U82621; AAB58154.1;
  KM Lipopolysaccharide biosynthesis: Transferase; Acyltransferase;
  KM Transmembrane; Inner membrane; Plasmid.
  FT NON_TER 1
  FT TRANSMEM 48 68 POTENTIAL.
  SQ SEQUENCE 231 AA; 26754 MW; 99350DB0A499D854 CRC64;

```

alignment_scores:

Quality: 120.50 Length: 249
 Ratio: 0.996 Gaps: 10
 Percent Similarity: 48.594 Percent Identity: 22.088

alignment_block:

us-09-303-518d-571 x MSBB_SHIFL ..
 Align seg 1/1 to: MSBB_SHIFL from: 1 to: 231

```

217 GTCAAGCCGTTTTTGGCGAAGCGCAAAATGCGGTTTGGAACTGCCCC 266
    ::::::::::::::::::::
1 ValAspAsnMetPheAlaThrAlaLeuGlnSerIleValMetMetLeu1 17
267 GCGCTTTTCAAAAACCGCAAGACATCGAACAATGTTCAAGCGGTA. 315
    ::::::::::::::::::::
17 uLeuAlaIleArgGlyProGluLysPheGln.....LysArgValP 31
316 ....CAGCGCTGGAAACAGCTGCACAGCGCTTGGACAAAGCGGCAAGG 360
    ::::::::::::::::::::
31 hetrPylsGlyLeuGluIleLeuGluGluIleArgHisAsnAsnArgAsn 47
361 CTGCTGTTCATCACGCGCGCATCGACAGCTAGCATTTGGCGGCGGCTA 410
    ::::::::::::::::::::
48 ValIlePheLeuValProHisGlyTrpSerValAspIleProAlaMetLe 64
411 CATCAGCCAGCAAGCTTCCTCCACCTGACCGCGCATGACAAAGCGCGCA 460
    ::::::::::::::::::::
64 uLeuAlaAlaGln...GlyGluLysMetAlaAlaMetPheHisGlnGlnA 80
461 AATCAAGCGATAGACAAATATCATGCAAGCGCGGAGGTCGCGCGA 510
    ::::::::::::::::::::
80 rGAsnProValIleAspTyrValIleProAsnSerValArg...ArgLysPhe 95
511 GCGAAACCGCGCGCCAGCGCATACAAAGGTCGAACAACATCATCAAGCG 560
    ::::::::::::::::::::
96 GluGlyArgLeuHisSerArgGluAspGlyIleLysProPheIleGlnSe 112
561 CTGCGCGCGGCGGCGAGCAACATCATCTGCC.....GACCAGCTCC 604
    ::::::::::::::::::::
112 rValArgGlnIleGlyTrpGlyTyrTyrLeuProAspGlnAspHisGlyP 129
605 CTTCGCGCAGGAAGCGCGCGCTGCGCGGATTTTTCGCGCAACCT 654
    ::::::::::::::::::::
129 rGluTyrSerGlu.....PheAlaAspPheAlaThrTyr 141
655 GCATACACCATGACACGCGCGCAAAATTTGGCACAGCTCAAGCGCTGA 704
    ::::::::::::::::::::
142 LysAlaThrLeuProIleIleGlyArgLeuMetAsnIle..... 154
705 AACCTGTTTTTGTGTCGAAGCGCTGCGCGAGCAAGCGCTTCGTCT 754
    ::::::::::::::::::::
155 .....SerGlnAlaMetIleI 160
755 TGCACATCGCGCGCGTCGAAGGGAATTGAAGCGCAACAAAGCCGAC... 801
    ::::::::::::::::::::
160 LeProLeuPheProValIleTyrAspGlu.....LysLysHisPhe 172
802 .....GATGCCGCGCTG..... 813
173 LeuThrIleGluValArgProPheMetAspAlaCysIleAlaSerAlaAs 189
814 .....TTCACCGCAATTCGCAATATGGATAC 841
    ::::::::::::::::::::
189 rAsnLysMetIleAlaArgGlnMetAsnLysThrValGluIleLeuValG 206
842 GCGGTTTCCGACGAGTATCTGTTATGTACAAACCGTATAAAGC 888
    ::::::::::::::::::::
206 LysHisProGluGlnTyrIleTrpValIleuLysLeuLeuLysThr 221
seq_name: SwissProt_40:NFX1_HUMAN

```

seq_documentation_block:

ID NFX1_HUMAN STANDARD: PRT: 1104 AA.
 AC 012986:
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Transcriptional repressor NF-X1 (Nuclear transcription factor, X box-
 binding, 1).
 GN NFX1.
 OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=95053707; PubMed=7964459;
 RA Song Z., Krishna S., Thanos D., Strominger J.L., Ono S.J.;
 RT "A novel cysteine-rich sequence-specific DNA-binding protein
 RT interacts with the conserved X-box motif of the human major
 RT histocompatibility complex class II genes via a repeated Cys-His
 RT domain and functions as a transcriptional repressor."
 RL J. Exp. Med. 180:1763-1774(1994).
 CC -1- EXP. Med. 180:1763-1774(1994).
 CC MOTIF OF CLASS II MHC GENES. MAY PLAY AN IMPORTANT ROLE IN
 CC REGULATING THE DURATION OF AN INFLAMMATORY RESPONSE BY LIMITING
 CC THE PERIOD IN WHICH CLASS II MHC MOLECULES ARE INDUCED BY
 CC INTERFERON-GAMMA.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- INDUCTION: BY INTERFERON GAMMA.
 CC -1- SIMILARITY: TO D.MELANOGASTER SHUTTLE CRAFT PROTEIN (STC) AND
 CC YEAST YNL023C.
 CC -1- SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC The European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: U15306; AAA69517.1;
 DR HSSP: P02876; ZMGC.
 DR MIM: 603255;
 DR InterPro: IPR001965; PHD.
 DR InterPro: IPR001374; R3H.
 DR InterPro: IPR000967; ZNF_NFX1.
 DR InterPro: IPR001841; Znf_ring.
 DR Pfam: PF01424; R3H; 1.
 DR Pfam: PF01422; zf-NFX1; 8.
 DR SMART: SM00393; R3H; 1.
 DR SMART: SM00184; RING; 1.
 DR SMART: SM00438; ZNF_NFX; 9.
 DR PROSITE: PS00518; ZF_RING_1; FALSE_NEG.
 DR PROSITE: PS50089; ZF_RING_2; 1.
 DR Transcription regulation; Repressor; DNA-binding; Nuclear protein;
 KW Repeat; Zinc-finger.
 KW Repeat; Zinc-finger.
 FT DOMAIN 342 393 RING-TYPE.
 FT REPEAT 424 876 7 X APPROXIMATE REPEATS, CYS-RICH.
 FT REPEAT 424 463 1.
 FT REPEAT 480 517 2.
 FT REPEAT 541 580 3.
 FT REPEAT 606 647 4.
 FT REPEAT 695 734 5.
 FT REPEAT 806 842 6.
 FT REPEAT 843 876 7.
 SO SEQUENCE 1104 AA; 123149 MW; 36FEC4292F78130C CRC64;

alignment_scores:
 Quality: 119.00 Length: 374
 Ratio: 0.937 Gaps: 17
 Percent Similarity: 33.957 Percent Identity: 19.786

alignment block:
 US-09-303-518D-571 x NFX1_HUMAN ..

Align seg 1/1 to: NFX1_HUMAN from: 1 to: 1104

14 AATTGAGGCTGTTTCCCTTTCGACGAGG...CCATGCAACATCCGCT... 58
 ||| ||| ::|||::||| ||||| ::|||
 553 AsnHisThrCysSerGlnValCysHisProGlnProCysGlnGlnCysPrr 569

59TGACGGCCCTGCTCAATGCGCCCTGCTGCTGCTGCTGCTCTCT 104
 ||| ||| ::|||
 569 oATgLeuProGlnLeuValArgCysCysProCysGlnGlnProLeuS 586
 105 GCACACGCTGGGAAACCGGCTCGACATCTGGCGTTTACCTTTAAAG 154
 ||| ||| ||| |||
 586 eArgLLeuLeuGlnLeuGlnLysSerSerSer.....Arg 596
 155 AAGACCGCGCGCGCATGCTGCGCAATATGCGCGCGGTTTGAACCC 204
 ||| ||| ||| |||
 597 LysThrCysMetLysProValProSerCysGly..... 607
 205 GACACGCGACGCTCAAGCGCTTTTTCGGAACGGCAAAATGCGGTT 254
 607 607
 255 GCAACTGCCCCCGCTTTTCAAAAACGGAAACATGCAACATG 304
 ||| |||
 608LysValCys 610
 305 TCAAGCGGTACACGCGTGGG.....AACGCTGCGACG 339
 ||| |||
 611 GlyLysProLeuProCysGlySerLeuAspPheLLeuHisThrCysGly 627
 340 GCTTTGACAGGCGGAAAGGCTGCTTTCATCAAGCGCGCATCGGCGAG 389
 ||| ||| ||| ||| ||| ||| |||
 627 sLeu...CysHisGlnGlnLysAspCysGlyProValSerArgThrVal. 642
 390 CTATGATTTGGCGGACGCTATCATGACGACGCTTCCTTCACCTGA 439
 ||| ||| ||| ||| ||| ||| |||
 643LysSerCysArgCysSerPheArgThrLysGlu 653
 440 CCGCATGTACAAAGCGCGCAAA.....TCAAA 468
 ||| ||| ||| ||| ||| ||| |||
 654 LeuProCysThrSerLeuLysSerGlnAspAlaThrPheMetCysAsp 670
 469 GCGATGACAA.....TCATGCAAGCGGCGAG 497
 ||| ||| ||| ||| ||| ||| |||
 670 sArgCysAsnLysLysArgLysCysGlnArgHisLysCysAsnGluIleC 687
 498 GGTGCGCGGCAAGGCAAAACCGCGCCA..... 526
 ||| ||| ||| ||| ||| ||| |||
 687 yscysValAspLysGlnHisLysCysProLeuAsnCysGlnArgLysLeu 703
 527CGGCATACAAAGGCGCAACAA 549
 ||| ||| ||| ||| ||| ||| |||
 704 ArgCysGlnLeuHisArgCysGlnGluProCysHisArgGlnLysCysG 720
 550 ATCA..... 553
 ||| ||| ||| ||| ||| ||| |||
 720 nThrCysTrpGlnAlaSerPheAspGlnLeuThrCysHisCysGlnAla 737
 554TCAGGCC 562
 ||| ||| ||| ||| ||| ||| |||
 737 eValLLeuTrpProValProCysGlnThrArgProGluCysThr 753
 563TCGCGCGCGGCGAGGCA..... 580
 ||| ||| ||| ||| ||| ||| |||
 754 GlnThrCysAlaArgValHisGlnCysAspHisProValThrHisSerG 770
 581CCATCATCTGCGCGACCGACGCTCTTCGCGAGAGGCGG 623
 ||| ||| ||| ||| ||| ||| |||
 770 yHisSerGlnGlnLysCysProProCysThrPheLeuThrLys..... 785
 624 CGGCGTGTGGCGGATTTTTCGGAACCGCAACACATGACACTGG 673
 ||| ||| ||| ||| ||| ||| |||
 786Trp 786
 674 CGGCAAAATTTGGCACACGTCAAAGCGGTGAAACCTGTGTTT..... 715
 ||| ||| ||| ||| ||| ||| |||
 787 CysMetCysLysHisGlnLysPheArgSerAsnLleProCysHisLeuValAs 803
 716TCTGCTGCAAGCGCTGCGCG..... 736

```

803 p15sercysglyleuprocysserAlaThrLeuprocyslymethisl 820
737 .....ACGACAAAGCTTCGCTTCACATCCGCCCGCTCC 772
820 yscysglinargleucysglisylsleuValaspGlnprCys 836.
773 AAGGGAATTGACGCCAACAAG.....CCGACGATGCC 807
837 LysGlnprCysThrThrProArgAlaAspCyslysisProcysMetAl 853
808 GCCGTGTTCACCGCAATACCG 829
853 aprocysHisThrSerSerPro 860

seq_name: SwissProt_40:HKR3_HUMAN

seq_documentation_block:
ID HKR3_HUMAN STANDARD; PRT; 688 AA.
AC P10074;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Kruppel-related zinc finger protein 3 (HKR3 protein).
GN HKR3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
[1]
RP SEQUENCE FROM N.A.
RA MEDLINE=95059073; PubMed=7969177;
RA Sugawara M., Scholl T., Ponath P.D., Strominger J.L.;
RT "A factor that regulates the class II major histocompatibility
RT complex gene DPA is a member of a subfamily of zinc finger proteins
RT that includes a Drosophila developmental control protein.";
RL Mol. Cell. Biol. 14:8438-8450(1994).
[2]
RP SEQUENCE FROM N.A.
RA MEDLINE=98177483; PubMed=9516840;
RA Maris J.M., Jensen J., Sulman E.P., Beltinger C.P., Allen C.,
RA Biagel J.A., Brodeur G.M., White P.S.;
RT "Human Kruppel-related 3 (HKR3): a candidate for the 1p36
RT neuroblastoma tumor suppressor gene?";
RL Eur. J. Cancer 33:1991-1996(1997).
[3]
RP SEQUENCE OF 461-488 FROM N.A.
RA MEDLINE=89096896; PubMed=2850480;
RA Ruppert J.M., Kinzler K.W., Wong A.J., Bigner S.H., Kao F.T.,
RA Law M.L., Seunasz H.N., O'Brien S.J., Vogelstein B.;
RT "The G1I-Kruppel family of human genes.";
RL Mol. Cell. Biol. 8:3104-3113(1988).
[4]
RP FUNCTION: BINDS TO AND REGULATES THE J AND/OR S ELEMENTS IN MHC II
RP PROMOTER.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
CC FINGER PROTEINS.
CC -1- SIMILARITY: CONTAINS 1 BTB/POZ DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; L16896; AAA65124.1; -
CC EMBL; U45325; AAB08973.1; -
CC DR EMBL; U45324; AAB08973.1; JOINED.
CC DR EMBL; M20677; AAA35989.1; -
CC DR PTR; E31201; E31201.
CC HSSP; P08047; ISP2.

```

```

DR MIM; 165270; -
DR InterPro; IPR000210; BTB_POZ.
DR InterPro; IPR000822; ZnF-C2H2.
DR Pfam; PF00651; BTB; 1.
DR Pfam; PF00096; ZF-C2H2; 11.
DR PRINTS; PR00048; ZINCFINGER.
DR SMART; SM00225; BTB; 1.
DR SMART; SM00355; ZnF-C2H2; 11.
DR PROSITE; PS50097; BTB; 1.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 9.
DR PROSITE; PS50157; ZINC_FINGER_C2H2_2; 11.
DR Transcription regulation; Activator; Zinc-finger; DNA-binding;
KW Repeat; Metal-binding; Nuclear protein.
FT DOMAIN 26 89
FT ZINC FINGERS.
FT ZN_FING 291 600
FT ZN_FING 291 313
FT ZN_FING 319 342
FT ZN_FING 350 372
FT ZN_FING 378 401
FT ZN_FING 407 430
FT ZN_FING 436 459
FT ZN_FING 465 487
FT ZN_FING 493 515
FT ZN_FING 521 544
FT ZN_FING 550 572
FT ZN_FING 578 600
FT CONFLICT 201 201
FT CONFLICT 244 244
FT CONFLICT 350 351
FT CONFLICT 607 607
SQ SEQUENCE 688 AA; 77054 MW; EBECCE3D6CB524 CRC64;

```

```

alignment_scores:
  Quality: 117.50
  Ratio: 1.078
  Percent Similarity: 45.228
  Percent Identity: 25.311

```

alignment_block:

US-09-303-518D-571 x HKR3_HUMAN

Align seg 1/1 to: HKR3_HUMAN from: 1 to: 688

```

84 CCGCTGCTGCGCTTCCTGCTGCACAGCTGGGAACCGCT..... 125
133 ProAlaSerGlnAsnValAsnSerHisValIsglnProAlaGlyLeuG1 149
126 .....CGACATCTGCGCTTTTAACTTTAAAGAGAGAC 159
149 uGlUGlUGlValIserArgThrLeuGlyLeuValProArgAspGlnIup 166
160 CGCGCGGCGCATCGTCGCCCATATCGCGAGCGGCTTGAACCCGAC 209
166 roHArgGlySerHisSerProGlnArgProGlnLeuHisSerProAlaGln 182
210 GCACAGCGGTCAAGCCGCTTTTGGGAACGCAAAATGCGGTTTGGAC 259
183 SerGlnGlyProSerSerLeuGlyGlyLeuValGlnAlaLeuGlySpr 199
260 TTGGCCCCCGGCTTTTCAAAAACCGGAGAGAC..... 291
199 oCyS.ProlauGluAspLysLysProGlnLysPcysLysValPropArg 215
292 ...ATCGAAACATGTTCAAGCGGTACAGGC.....TGGA 326
216 ProlauGlnAlaGlnGlyAlaGlnLeuGlnGlySerAsnGluTrpG1 232
327 ACACGTGCACAGCTTTGGACAAGGCGAAGGCGTGTTCATCAGCG 376
232 uValValIValGlnValGlnLysPaspGlyasp..... 242
377 CGCACATCGCAGCTACGATTTGGCGGAGCGTACATCAGCCAGAGCTT 426

```

```

243 ..... GlyaspTyrMetSerGluPro... 249
427 CCGTTCACCTGACCGCATGTACAGCCCGCCGAAATCAAGCATAGA 476
    |||::: |||::: |||::: |||::: |||:::
250 ..... GluAlaVal.LeuThrArgArgLysSer..... 258
477 CAAATCATGACGAGCGGCGAGGTGCGCGCAAAAGCAACCGCCCA 526
    ::||| |||::: |||::: |||::: |||:::
259 ..... AsnValIleArgLysProCysAlaIleGlu..... 268
527 CCGCATCATCAGGGGTCAACAAATCATCAAGCGCGCGCGCGAG 576
    |||::: |||::: |||::: |||::: |||:::
269 ProAlaLeuSerIleArgLysLeuAlaIleGluProAlaGluAsnArgLys 285
577 G.....CAACCATCATCTGCGCGCACGATGCTTCTCCGACAGA 617
    :|||::: |||::: |||::: |||::: |||:::
285 sglThrAlaValProValGluCysProThr..... 295
618 AGCGGCGCGCGCTGTGGCGGATTTTTCGCAACGTCATACACATGA 667
    |||::: |||::: |||::: |||::: |||:::
296 ..... CysHisLysLysPheLeuSerLysTyrTyrLeuLysVal 308
668 CACTGCGCGCAAAATTTGGCACACGTCAAGCGCTGAAACCCCTGTTTC 717
    |||::: |||::: |||::: |||::: |||:::
309 HisAsnArgLys.....HisThrGlyLys.....LysProPhe..... 319
718 TGCTGCGAAGCGCTGCCCG 736
320 .....GluCysPro 322
seq_name: SwissProt_40:DBPA_RAT
seq_documentation_block:
ID DBPA_RAT STANDARD; PRT; 361 AA.
AC Q62764; Q62764;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE DNA-binding protein A (Cold shock domain protein A) (Muscle Y-box
DE protein YB2) (Y-box binding protein-A) (RYB-A).
GN CSDA OR DBPA OR YB2
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY;
RA Goldman D., Gao J., Burmeister M., Sapru M.;
RT "Characterization of muscle Y-box proteins that bind the mChR
RT delta subunit promoter.";
RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=MISTAR; TISSUE=Liver;
RX MEDLINE=94301785; PubMed=8029009;
RA Ito K., Tsutsumi K., Kuzumaki T., Gomez P.F., Otsu K., Ishikawa K.;
RT "A novel growth-inducible gene that encodes a protein with a
RT conserved cold-shock domain.";
RL Nucleic Acids Res. 22:2036-2041(1994).
CC -1 SUBCELLULAR LOCATION: Nucleat.
CC -1 ALTERNATIVE PRODUCTS: AT LEAST 2 ISOFORMS; 1 (SHOWN HERE) AND 2;
CC -1 ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1 TISSUE SPECIFICITY: ABUNDANT IN THE SKELETAL MUSCLE, SPLEEN, AND
CC FETAL LIVER.
CC -1 SIMILARITY: BELONGS TO THE COLD-SHOCK DOMAIN (CSD) FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb.ch/announce/

```

```

CC or send an email to license@isb.ch).
CC -----
DR EMBL; U22893; AAB60520.1; -.
DR EMBL; D28557; BAA05907.1; -.
DR HSSP; P15277; IMC.
DR InterPro; IPR002059; Cold_shock.
DR Pfam; PF00313; CSD; 1.
DR PRINTS; PR00050; COLDSHOCK.
DR PRODOM; PD00621; Cold_shock; 1.
DR SMART; SM00357; GSP; 1.
DR PROSITE; PS00352; COLD_SHOCK; 1.
KW Transcription regulation; Repressor; DNA-binding; Nuclear protein;
KW Alternative splicing.
FT DOMAIN 85 149 CSD.
FT VARSPIC 184 252 MISSING (IN ISOFORM 2).
FT CONFLICT 14 14 L->H (IN REF. 2).
FT CONFLICT 52 74 SPGGDADGAPASAPAGSEDA -> APARAPRARPGLI
FT CONFLICT SPRKKRR (IN REF. 2).
FT PR -> HV (IN REF. 2).
FT SEQUENCE 361 AA; 38851 MW; C6799D5A3DA5C3F3 CRC64;
SQ

```

```

alignment_scores:
  Quality: 110.50      Length: 201
  Ratio: 1.315        Gaps: 8
  Percent Similarity: 41.791      Percent Identity: 24.876

```

```

alignment_block:
US-09-303-518D-571 x DBPA_RAT ..

```

```

Align seg 1/1 to: DBPA_RAT from: 1 to: 361

```

```

147 TTAAAGAGACCGCGCGCATGCTCGCATATAGCG..... 185
    ::|||::: |||::: |||::: |||::: |||:::
173 TyrTyrGlyArgArgArgGlyProProArgAsnTyrAlaGlyGluGlu 189
186 .....GCAGCGGTTTGAACCCGACACAGCAGAGC 216
189 uGluGluGlySerGlySerSerGluGluGlyPheGluProProAlaIleAsp 206
217 GTCAAAGCCGTTTTCGGAACGCGCAAAATGCGG..... 251
    |||::: |||::: |||::: |||::: |||:::
206 LysIlePheSerGlyAlaArgAsnGlnLeuArgArgProGlnTyrArgPro 222
252 .....TTGG 256
223 ProTyrArgGlnArgPheProProTyrHisValGlyGlnThrPheAs 239
257 AACTTGCCCGCGCGCTTTTCAAAAACCGAAGACATGCAACATGTTTC 306
    |||::: |||::: |||::: |||::: |||:::
239 PArgArgSerArgValPhePro.....HisProAsnArgMetG 252
307 AAACCGGTACACGCGTGGGAACAGCTGACAGCGCTTTGGACAAGGCGA 356
    |||::: |||::: |||::: |||::: |||:::
252 IlnIleGly...GluIleGlyGluMetLysAspGlyValProGluGlyAla 267
357 AGGCGTGTTCATCAACGCGCGCATCGGCGAGCTACGATTGGCGGAC 406
    ::|||::: |||::: |||::: |||::: |||:::
268 GlnLeuGlnValHisArgAsnProThrTyrArgProArgPheArgGly 284
407 GCTACATCAGCCACAGCTTCGTTCCACCTGACCGGCATGTACAAACCG 456
    ::|||::: |||::: |||::: |||::: |||:::
284 YProAlaArgPro.....ArgProAlaProAlaI 294
457 CCGAAATCAAGGATGACAAATCATGACGCGGCGAGGTGCGCGG 506
    ::|||::: |||::: |||::: |||::: |||:::
294 IeGlyIleAlaGlnAspLysGluAsnGlnIleAlaIleAsnGlyProAsn 310
507 CAAGAGCAAAACCGCGCGCGCGCATACAAAGGGGTCAACAAATCATCA 556
    |||::: |||::: |||::: |||::: |||:::
311 Gln..... 311
557 AGGCCCTGCGCGCGGCGGCGAGGCAACATCATCT.....GCC 594

```

```

312 .....ProSerAlaArgGlyPheArgArgProTyrAsnTyrArgArg 327
595 GACCAGCTCCCTCTCCGCA.....GGAAGCGCGCGGTGTG 632
327 rprProArgProLeuAsnAlaValSerGlnAspGlyLysGlnThrLysAla 343
633 GGC 635
344 Gly 344

```

seq_name: SwissProt_40:VE2_HPV19

seq_documentation_block:

```

ID VE2_HPV19 STANDARD; PRT; 493 AA.
AC P36786;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Regulatory protein E2.
GN E2.
OS Human papillomavirus type 19.
OC Viruses; dsDNA viruses, no RNA stage; Papillomaviridae;
OC Papillomavirus.
OX NCBI_TaxID=10608;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=94265501; PubMed=8205838;
RT Delius H., Hofmann B.;
RT "Primer-directed sequencing of human papillomavirus types.";
RT Curr. Top. Microbiol. Immunol. 186:13-31(1994).
CC -1- FUNCTION: E2 REGULATES VIRAL TRANSCRIPTION AND DNA REPLICATION.
CC IT BINDS TO THE E2E RESPONSE ELEMENT (5'-ACCNANNNGGT-3') PRESENT
CC IN MULTIPLE COPIES IN THE REGULATORY REGION. IT CAN EITHER
CC ACTIVATE OR REPRESS TRANSCRIPTION DEPENDENT OF E2E'S POSITION
CC WITH REGARDS TO PROXIMAL PROMOTER ELEMENTS. REPRESSION OCCURS
CC BY STERICALLY HINDERING THE ASSEMBLY OF THE TRANSCRIPTION
CC INITIATION COMPLEX. THE E1-E2 COMPLEX BINDS TO THE ORIGIN OF DNA
CC REPLICATION.
CC -1- SUBUNIT: BINDS DNA AS A DIMER.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X74470; CA52521.1; -
DR PIR: S36488; S36488.
DR HSSP: P03122; ZBOP.
DR InterPro: IPR001866; E2_N.
DR InterPro: IPR000427; Early2_C.
DR Pfam: PF00511; E2_C; 1.
DR Pfam: PF00508; E2_N; 1.
DR ProDom: PD000672; Early2_C; 1.
DR ProDom: PD000678; E2_N; 1.
DR Early protein; transcription regulation; Activator; DNA-binding;
KW Trans-acting factor; DNA replication; Repressor; Nuclear protein.
SQ SEQUENCE 493 AA; 55740 MW; DFE8F316FC66C782 CRC64;

```

alignment_scores:

```

Quality: 110.50 Length: 244
Ratio: 0.929 Gaps: 8
Percent Similarity: 48.770 Percent Identity: 23.361

```

alignment_block:

US-09-303-518d-571 x VE2_HPV19

Align seg 1/1 to: VE2_HPV19 from: 1 to: 493

```

200 ACCCCGACACGACGACGCTCAAGCCGTTTTCGGAACGGCAAAATGC 249
224 ThrAspSerAlaSerArgLeuSerProThrAlaSerArgLysGlnSerG1 240
250 GGTTCGTAACCTTGCCCGCCGCTTTTCAAAAACCGGAACGATCGAA. 298
240 ngInThrAsnThrLysGlyArgArgTyrGlyLysArgProSerSerArgT 257
299 .....CATGTTCAAGCGGTACAGCGGCGGGAACACGCGC 334
257 hrArgArgGlnThrGlnThrArgGlnLysArgSerArgSerLysSerLys 273
335 ACCAGCGTTTGACCAAGGCGGAGGCGTGTTCATCAGCCGACATC 384
274 SerArgSerArgSerArgSerArgSerArgSerArgSerArgSerArg 290
385 GGCACCTTCAGATTGGCGGAGCGTACATCAGCCAGCAGCTCCGTTCCA 434
290 SerLysSerArgArgLysAlaSerThrThrArgGlyArgGlyArgGlys 307
435 CCGTACCGCCATGTACAAAGCCGCAAAATCAAGCGATGACCAAAATCA 484
307 er...ProThrAlaThrSerAspGlnSerArgSerArgProSerAlaThr 322
485 TGCAAGCGGCGGAGGCGGCGGCAAAAGCAAAACCGCGCCACCGGCATA 534
323 .SerSerThrThrSerLeuArgSerArgLysSerArgValGlyArgS 339
535 CAAGGCGCAACCAATCATCAAGCGCGCGCGGCGGAGCAACCAT 584
339 erArgGlyGlyArgSerArgValGlyArgSerArgLysArgLysArg 355
585 CATCTGCCCGACCGACGCTCCCTCCGAGGAGGCGGCGGTGTGG 634
356 SerArgLysSerProSerProThrAsnThrLysArgSerArgGlnSe 372
635 CGGATTTTTCG. ....CAACCTGCA 657
372 rGlySerSerArgLeuHisGlyValSerAlaAspAlaValGlyThrSer 389
658 TACACCATGACATGCG.....GGCAAAATGCG.....ACAGCT 692
389 alHisThrValSerGlyArgAsnThrGlyArgLeuGlyArgLeuLeu 405
693 CAAGGCGTGAACCCGCTGTTTTCGTCGCA.....ACGCTGC 733
406 GluAlaLeuAspProValAlaLeuValArgGlyLysProAsnThrLe 422
734 CCGACGACACGCTGCTGTTCGACATCCGCCCGCTCAAGGGGAATTG 783
422 uArgSerPheArgAsnArgAlaLysHis..... 431
784 AACGGCAACAAAGCCACGATCCGCGCTTCAACCGCAATACCGAATA 833
432 .....MetTyrArgGly 435
834 TTGATACGCGCTTTCGACGACGATCT 863
436 LeuPheSerSerPheSerThrAlaTyrSer 445

```

seq_name: SwissProt_40:MP5A_LOLPR

seq_documentation_block:

```

ID MP5A_LOLPR STANDARD; PRT; 308 AA.
AC Q40240;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Major pollen allergen Lol p 5a precursor (Lol p Va) (Lol p Ib).
GN LOLPB.
OS Lolium perenne (Perennial ryegrass).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

```

```

638 ATTTTTCGGC.AAACCGCATACACCATGACACTGGCGGCAAAATTGGC 686
152 ValHisAlaValLysProAlaThrGluGluValLeuAlaAlaLysIlePr 168
687 ACACGTCAAG.....GCGTGAAGAACCC 709
168 o.ThgIgcTgIgcGlnIleValAlaPylsIleAspAlaAlaPheLysIle 184
710 TGTTCCTGCTGCGAACGCCGTCGCCGAG..GACAGCGCTGGTGT.. 754
185 AlaAlaThrAlaAlaAsnAlaAlaProThrAsnAspLysPheThrValPh 201
755 .....TGCACA..... 760
201 egluSerAlaPheAsnLysAlaLeuAsnGluLysThrGlyGlyAlaMet 218
761 .....TCCGCCCGCTCCAGGGGAATTGACGCGCAACAAAG 796
218 rprThrSerSerProProSerIleArgPro....SerSerArg 232
797 CCACGATCGCGCGGTTCACCGCATACCGAATTGGATACCGCGT 846
233 ProThrProPro....SerProAlaAlaProGluValLysThrAlaVal 248
847 TTTTCGACGCGAGTATCTGTTATGTACACCGCTATA 883
248 lPheGluAlaAlaLeuThrLysAlaIleThrAlaMet 260

seqname: Swissprot_40:YFHA_BORPE

seq_documentation_block:
ID YFHA_BORPE STANDARD: PRT; 320 AA.
AC P33445;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 33.8 kDa protein in FNA3 3'region (ORFA).
OS Bordetella pertussis.
OC Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;
OC Bordetella.
OX NCBL_Taxid=520;
RN [1]
RP SEQUENCE FROM N.A.
RA Williams R.J.L.;
RL Submitted (JUN-1993) to the EMBL/GenBank/DBJ databases.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its use
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
DR EMBL; X64876; CAA46093.1; -
KW Hypothetical protein.
SQ SEQUENCE 320 AA; 35781 MW; 185C54B11DEF2D73 CRC64;

alignment_scores:
Quality: 108.50 Length: 261
Ratio: 0.882 Gaps: 13
Percent Similarity: 47.126 Percent Identity: 26.437

alignment_block:
US-09-303-518D-571 x YFHA_BORPE ..

Align seq 1/1 to: YFHA_BORPE from: 1 to: 320

171 CGTCGCCAATATGCGCGAGCGGTTTGAAACCCGACACGACAGACGCTCA 220
|||||: ::||| ||| ||| ||| |||
3 ArgGluGlnLeuGlyAlaGlyGlnAspGlnAspGlnAspHisAla...ValGI 18

```

```

221 AAGCGCGTTTGGCGGAAGCGCAAAATGCGGTTTGGACCTTGGCCCGCG 270
18 :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
271 TTTTCAAAAAACCGAAGACATCGAAGACATGTCTTCAACCGGTACACG 320
35 :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
321 CTGGGACACGTGCAGACGCTTTGGCAAGGCGGAAGGCTGTCTTCA 370
52 :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
371 TCAAGCGGCACATCGCGACTTACA.....TTGGCGCGA 405
65 :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
371 :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
406 CGGTACATCAAGCGACAGCTCCGTT..... 431
82 :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
432 :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
99 :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
452 AAGCGCGCAAAATCAAGCGATAGCAAAATCATCGACGCGCGAGGTG 502
115 :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
115 :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
502 CGCGCAAAAGCGCAAGCGCGCGCGCATACAGGCGCAACAAAT 551
132 :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
552 :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
145 :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
585 :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
161 :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
628 :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
178 :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
675 :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
195 :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
725 :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
208 :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
753 :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
223 :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
792 :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
236 :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
seq_name: SwissProt.40:VE2_HPV47
seq_documentation_block
ID VE2_HPV47 STANDARD; PRT; 506 AA.
AC P22420;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Regulatory protein E2.
GN E2.
OS Human papillomavirus type 47.
OC Viruses; dsDNA viruses, no RNA stage; Papillomaviridae;
Papillomavirus.

```

```

XN  NCBI_TaxID=10594:
RN  [1]
RP  SEQUENCE FROM N.A.
RX  MEDLINE=90281611; PubMed=2162112;
RA  Kiyono T., Adachi A., Ishibashi M.;
RT  "Genome organization and taxonomic position of human papillomavirus
RL  type 47 inferred from its DNA sequence.";
    viology 177:401-405(1990).
CC  -I- FUNCTION: E2 REGULATES VIRAL TRANSCRIPTION AND DNA REPLICATION.
CC      IT BINDS TO THE E2RE RESPONSE ELEMENT (5'-ACCGNNNNNGGT-3') PRESENT
CC      IN MULTIPLE COPIES IN THE REGULATORY REGION. IT CAN EITHER
CC      ACTIVATE OR REPRESS TRANSCRIPTION DEPENDING OF E2RE'S POSITION
CC      WITH REGARDS TO PROXIMAL PROMOTER ELEMENTS. REPRESSION OCCURS
CC      BY STERICALLY HINDERING THE ASSEMBLY OF THE TRANSCRIPTION
CC      INITIATION COMPLEX. THE E1-E2 COMPLEX BINDS TO THE ORIGIN OF DNA
CC      REPLICATION.
CC  -I- SUBUNIT: BINDS DNA AS A DIMER.
CC  -I- SUBCELLULAR LOCATION: Nucleat.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on way
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL; M32305; AAA46679.1; -.
DR  PIR; D35324; W2ML47.
DR  HSSP; P03122; 2BOP.
DR  InterPro; IPR001866; E2_N.
DR  InterPro; IPR000427; Early2_C.
DR  Pfam; PF00511; E2_C; 1.
DR  Pfam; PF00508; E2_N; 1.
DR  ProDom; PD000672; Early2_C; 1.
DR  ProDom; PD000678; E2_N; 1.
DR  Early protein: Transcription regulation; Activator; DNA-binding;
KW  Trans-acting factor; DNA replication; Repressor; Nucleat protein.
SO  SEQUENCE 506 AA; 57478 MW; 92C37F4BF8725065 CRC64.

```

alignment_scores:	
Quality:	108.00
Ratio:	0.857
Percent Similarity:	47.191
Percent Identity:	23.070

alignment_block:

US-09-303-518D-571 x VE2_HPV47

Align seg 1/1 to: VE2_HPV47 from: 1 to: 506

```

44 CCATCAGACATCCCTGTGACCGCCCTGCTA.....AATGC 78
197 ProValThrSerSerThrProProGlySerProGlyGlnThrAspPr 213
79 CTCCTCCGCGCTGCGCTTTCCTCTGCACACCGCTGGGAACCGGCTCGG 128
213 oasPthSerSerIythrProthrThrThrAlaIalThrSphrs 230
129 ACATCTGCGCGCTTTTACCTTTTAAAGGACAGCCGCGGCATCTGCGCCA 178
230 erProAlaArgGlnSerIleAsnLysGlnSerGlnGlnThrLys 246
179 ATA.....TGGCGGACGAGCGGGCTTTGAAACCCGAGACGCAAGCGTCAAA 222
247 ArgArgGlyIyIyIyIyArgArg.....ProSerIerArgThrArg 260
223 GCGGTTTTCGGGAACGGCAAAATGCGGTTTGGAACTTGCCCCGGGTT 272
260 gProGlnThrHisGlnIyArgArg..... 267
273 TTTCAAAAAACCGGGAACATCTGAAACCATTTTCAAACCGGTACACGGCT 322
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

CC      VISCIO-ELASTIC PROPERTY OF WHEAT DOUGH.
CC      -1- SUBUNIT: DISULFIDE-BRIDGE LINKED AGGREGATES.
CC      -1- MISCELLANEOUS: GLUTENINS ARE CODED BY SEVERAL GENES ON EACH OF THE
CC      CHROMOSOMES OF WHEAT.
CC      -1- MISCELLANEOUS: THE MATURE PROTEIN IS CHARACTERIZED BY A LARGE
CC      NUMBER OF WELL PRESERVED REPEATS OF THE TWO MOTIFS: GOOPQGD AND
CC      GOOPRGQGQGYTPRS
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      DR      EMBL; X12928; CAA31395.1; .
CC      DR      PIR; S02262; S02262.
CC      DR      InterPro: IPR001419; Glutenin.
CC      KW      PRINTS; PR00210; GLUTENIN.
CC      FT      Signal storage protein; Repeat: Multigene family: signal.
CC      FT      SIGNAL          1         21
CC      FT      CHAIN           22        839
CC      FT                                     DKS.
CC      FT                                     REPEATS.
CC      FT      DOMAIN       131         801
CC      FT      SEQUENCE     839 AA; 89359 MW; 0F14E1106D552643 CRC64;
CC
CC      alignment_scores:
CC              Quality: 107.00          Length: 254
CC              Ratio: 0.843             Gaps: 14
CC      Percent Similarity: 50.000      Percent Identity: 23.228
CC
CC      alignment_block:
CC      US-09-303-518D-571 x GLT5_WHEAT ..
CC
CC      Align seg 1/1 to: GLT5_WHEAT from: 1 to: 839
CC
CC      63 CGCGCTGCATAATGCCCTCCCTGCTGTGGCGTTTCGTCGACAGC 112
CC      :::::::::::::::::::: ||| ::::::::::::::::::::
CC      175 GlnProlyGlnLeuGlnInPro.....AlaGlnI 185
CC
CC      113 TGGGAACCGCGCTCGG...ACATCTGGCGTTTACCCTTTTAAGSAGAC 159
CC      : :::::::::::::::::::: ::::||| ::::::::::::::::::::
CC      185 yGlnGlnProGlyGlnGlnGlnGlnGlnGlnProGlyGlnGlnGln 202
CC
CC      160 CG.....CGCGCGCATCGTGCCAA 179
CC      ||| :::::::::::::::::::: ::::| ::::::::::::::::::::
CC      202 roGlyTyrrProthrSerSerGlnLeuGlnProGlyGlnLeuGln 218
CC
CC      180 TATGCGGACGCG.....GGGTTGAACCCCGCACACGCGAGC..... 215
CC      ||| ||| :::::::::::::::::::: ::::::::::::::::::::
CC      219 ProAlaGlnGlnGlnGlnGlnGlnProGlyGlnAlaGlnGlnGln 235
CC
CC      216 .....GTCGAAGCGGTTTTGCGGAACGCGCAAAATGCGGTTGGAA 258
CC      :::::::::::::::::::: ::::::::::::::::::::
CC      223 nGlnProGlyGlnGlnGlnProGlyGlnGlnGln.....GlyG 250
CC
CC      229 CTTCGCCCGCGGTTTTTCAAAAAACCGGAAGACATGAAACAATGTCAA 308
CC      ||| :::::::::::::::::::: ::::| ::::::::::::::::::::
CC      250 InGlnProGlyGlnGlnGlnProGlyGlnGln.....Gln 262
CC
CC      309 AGCGGTACAGCGCTGGGACACGTGCACGACGCGGCTT..... 344
CC      ::::| | :::::::::::::::::::: ||| ::::::::::::::::::::
CC      263 GlnGlnGlnGlnLeuGlnGlnGlnGlnGlnGlnGlnGlnProThrserie 279
CC
CC      345 .....GGACAAGGCGAAGGCGGCTGTCATCAGCGCGCACATC 384
CC      :::::::::::::::::::: ||| ::::::::::::::::::::
CC      279 uGlnGlnSerGlyGlnGlnGln.....GlnProGlyTyrrProThrserleue 295
CC
CC      385 GGCAGCTACGATTGGCGGAGCGCTACTACGACGACGACTTCGTTCCA 434
CC      ::::| | | | | | | | | | | | | | | | | | | | | |

```



```

295 lnglnleuglnglnglnglnsergltYrTyrProthserProglngln 311
435 CCTGACGCCATGTACAGCCGCCGAAATCAACGATAGCAAAATCA 484
||| :|: ||| :|: ||| :|: ||| :|: ||| :|:
312 ProglnglnglnglnglnProglnglnglnglnProalnglnglng 328
485 TGCAGCGGCGAGAGGTGCGCGCAAGGCAACCGCGCCACCGGCA 534
:|: ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
328 nglnProglnglnglnglnglnglnglnglnglnglnglng 343
535 CAAGGGGTCAACAATCATCATCAAGC.....CCTGCGCGCGCGGCGAGGC 578
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
343 lnglnglnglnglnglnglnglnglnglnglnglnglnglnglng 359
579 AACCATCATCTCTGCCGACACGCTCTCTCTCGGAGGAGGCGGCGG 628
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
360 gltYrTyrProthserProglnglnglnglnglnglnglnglng 373
629 TGTGGCGGATTTTTCGCGCAACCTGCATACACCATGACACT..... 671
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
374 ....ProglYrTyr...ProthserSerGlnglnglngProthchlnserg 388
672 .....GGCGCAAAATTTGGCACACGTCGCAAGG 698
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
388 lnglnProglnglnglnglnglnglnglnglnglnglnglnglng 404
699 CGTGAACACCT 710
:|: ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
405 AlaglnglngPro 408

```

seq_name: SwissProt_40:SFR2_MOUSE

seq_documentation_block:

ID SFR2_MOUSE STANDARD: PRT: 221 AA.

AC 062093;

DT 15-DEC-1998 (Rel. 37, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DE 16-OCT-2001 (Rel. 40, Last annotation update)

DE Splicing factor, arginine/serine-rich 2 (Splicing factor SC35) (SC-35)

GN SFRS2 OR PR264.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxId=10090;

RN [1]

RP SEQUENCE FROM N.A.

RA MEDLINE=98447613; PubMed=9774382;

RA Yang L., Embree L.J., Tsai S., Hickstein D.D.;

RT "Oncoprotein TLS interacts with serine-arginine proteins involved in

RNA splicing.";

RL J. Biol. Chem. 273:27761-27764 (1998).

RP SEQUENCE OF 1-121 FROM N.A.

RA STRAIN=129/SV; TISSUE=Liver;

RA Galliard C., Perbal B.;

RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.

CC FUNCTION: NECESSARY FOR THE SPLICING OF PRE-mRNA. IT IS REQUIRED

CC FOR FORMATION OF THE EARLIEST ATP-DEPENDENT SPLICING COMPLEX AND

CC 3' SPLICING SITES DURING SPLICING ASSEMBLY. IT ALSO IS REQUIRED

CC FOR ATP-DEPENDENT INTERACTIONS OF BOTH U1 AND U2 SNRNPS WITH PRE-

CC mRNA (BY SIMILARITY).

CC SUBCELLULAR LOCATION: Nuclear.

CC -1- PTM: EXTENSIVELY PHOSPHORYLATED ON SERINE RESIDUES IN THE RS

CC DOMAIN (BY SIMILARITY).

CC -1- SIMILARITY: BELONGS TO THE SR FAMILY OF SPLICING FACTORS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

```

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF077858; AAC1000.1; -.
CC DR EMBL: X98511; CAAC67134.1; -.
CC DR HSSP: P09651; IUP1.
CC DR MGD: MGI:98284; Strs2.
CC DR InterPro: IPR000504; RRM.
CC DR Pfam: PF00076; Rrm; 1.
CC DR SMART: SM00360; RRM; 1.
CC DR PROSITE: PS50102; RRM; 1.
CC DR PROSITE: PS00030; RRM_RNP_1; 1.
CC KW Nuclear protein; RNA-binding; mRNA splicing; phosphorylation.
CC FT DOMAIN 14 92 RNA-BINDING (RRM).
CC FT DOMAIN 117 116 GLY-RICH (HINGE REGION).
CC FT DOMAIN 117 221 ARG/SER-RICH (RS DOMAIN).
CC SO SEQUENCE 221 AA; 25476 MW; 68121AC4D35714FA CRC64;

```

alignment_scores:

Quality:	Ratio:	Length:
105.50	1.045	211
Percent Similarity:	47.867	Gaps: 7
Percent Identity:		22.749

alignment_block:

US-09-303-518d-571 x SFR2_MOUSE

Align seg 1/1 to: SFR2_MOUSE from: 1 to: 221

```

258 ACPTGGCCCCCGCTTTTCAAAAACCGGAGACATGCAACATGTTCA 307
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
29 ThrleuArgArgValPheglutylTyrGlyArgValGlyAspValTyrI 45
308 AACCGGTACACGCGCTGGAAACAGCGTCACGACGCTTGA..... 347
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
45 eProArgAspArgTyrThrlyGlnSerArgGlyPheAlaPheValArgp 62
347 .....
62 heHisAspLysArgSpAlaGluAlaMetAspAlaMetAspValAla 78
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
348 .....CAAGGCGAAGGCGCTGCTTCATCAGCGCGCATCGGACCT 391
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
79 ValLeuAspGlyArgGluLeuArgValAlaMetAla..... 90
392 AGGATTTGGGCGGACGCTACATCAGCCAGAGCTTCGTCACCTGACC 441
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
91 ArgTyrGlyArg.....ProProAsps 98
442 GCCATGTACAAAGCCGCAAAATCAAGCGTATACAAATCATGACGC 491
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
98 eRHis...HisSerArgGlyProProProArgArgTyrGlyGly 113
492 GGGCAGGCGTGGCGGCAAGGCAAGCGCCGCGCATCAAGGG 541
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
114 GltYrGlyArgSerArgSerProArgArgArgArgSerArgse 130
542 TCAAAACATCATCAGGCGCTGGCGCGGCGGAGGACACCATCATCTG 591
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
130 ArgSerArgSerArgSerArgSerArgSerArgTyrArgS 147
592 CCGCAGCAGCTCTTCGCGAGAGGCGGCGGTGGCGGATTT 641
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
147 eRlySerArgSerArgTyrArgSerArgSerArg..... 158
642 TTTCGCAACCTGCATACCATGACATGCGGCAAAATTTGGACAGC 691
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
159 .....TCAAGGCGTGAACCTGTTTCTGCTGCAAGACCTGCCCA...C 738
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
167 gArgSerLysSerLysSerSerValSerArgSerArgSerA 184

```


FT	TRANSMEM	1666	1684
FT	DOMAIN	1685	1703
FT	TRANSMEM	1704	1723
FT	DOMAIN	1724	1795
FT	TRANSMEM	1796	1820
FT	DOMAIN	1821	2424
FT	DOMAIN	13	18
FT	DOMAIN	727	732
FT	DOMAIN	1004	1010
FT	DOMAIN	1012	1017
FT	DOMAIN	2219	2227
FT	DOMAIN	2242	2246
FT	DOMAIN	2288	2297
FT	DOMAIN	2298	2301
FT	DOMAIN	2372	2377
FT	DOMAIN	2411	2416
FT	DOMAIN	383	400
FT	SITE	318	318
FT	SITE	668	668
FT	SITE	1469	1469
FT	SITE	1765	1765
FT	MOD.RES	1831	1831
FT	CALBIND	1849	1860
FT	CARBOHYD	283	283
FT	CARBOHYD	1665	1665
FT	VARSPLIC	772	1051
FT	VARSPLIC	772	1120
FT	VARSPLIC	1857	1884
FT	VARSPLIC	2230	2273
FT	VARSPLIC	2274	2424
FT	VARIANT	419	419
FT	VARIANT	877	877
FT	VARIANT	1104	1104
FT	MUTAGEN	386	386
FT	MUTAGEN	389	389
FT	MUTAGEN	392	392
FT	MUTAGEN	400	400
SO	SEQUENCE	2424 AA:	273228 MM: F7CC4D0AB4B45604 CRC64;

```
alignment_scores:
  Quality: 105.50
  Ratio: 0.776
  Percent Similarity: 43.450
  Length: 313
  Gaps: 15
  Percent Identity: 22.364
```

```
alignment_block:
```

```

US-09-303-518D-5/1 x CCAA_RABIT  ..
Align seg 1/1  to: CCAA_RABIT  from: 1  to: 2422

```

```

10  CAGCCGCTGCTCCGCTTTGGAAACGGCAGTCACCTGGTGAAGCGCC 67
11  |||  :::::  :::::  :::::  |||  :::::  |||
2048  GlnlyThrGlyThrPrSerProGluAlaProProlAlaSpMetAl 206
68  TGTCAATGCGCTCCCT..CGTGGCTTTCTGTGACACGGTG 114
2064  AASPserGlnProLysProGlnSerValGluIleMetArgLysMetSerGlnA 2081
115  GGAACCGCGCTCGACATCTGGCGGTTTACTCTTTAAAGAGACCGGCG 164
116  ::  :::::  |||||  :::::  |||
2081  spGlyTyrSerAspSerGlnHisCysLeuProMetGlnGlyGlnIleLarg 2097
165  GCG.....CATGCTGGCAGATATGCG 184
11  |||  ::|||  :::::

```

[illegible]

DE (BASS1) (Protein C21orf50).
SON OR NREBP OR DBPS OR C21ORF50 OR KIAA1019.
OS Homo sapiens (Human).
CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
NCBI_TaxId=9606;
RX MEDLINE=21564202; PubMed=11707072;
RA Raymond A., Friedl M., Neergaard Henriksen C., Chapot F.,
Deutsch S., Ucla C., Rossier C., Lyle R., Guipponi M.,
Antonarakis S.E.;
RA "From PREDS and open reading frames to cDNA isolation: revisiting the
RT human chromosome 21 Transcription Map";
RL Genomics 78:46-54(2001).
[2]
RN SEQUENCE FROM N.A. (ISOFORM G).
RP TISSUE=Liver;
RC MEDLINE=21316479; PubMed=11306577;
RX Sun C.-T., Lo W.-Y., Wang I.-H., Lo Y.-H., Shou S.-R., Lai C.-K.,
Ting L.-P.;
RA "Transcription repression of human hepatitis B virus genes by negative
RT regulatory element-binding protein/SON";
RL J. Biol. Chem. 276:24059-24067(2001).
[3]
RN SEQUENCE OF 1-689 FROM N.A. (ISOFORM H).
RP TISSUE=Placenta;
RC Casadei R., Strippoli P., D'Addabbo P., Canaider S., Lenzi L.,
Vitali L., Giannone S., Carinci P., Zannotti M.;
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
[4]
RN SEQUENCE OF 1-130 FROM N.A.
RC TISSUE=Smooth muscle;
RA Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T., Matsumura K.,
Nakajima Y., Mizuno T., Morinaga M., Tanigami A., Fujiwara T., Ono T.,
Yamada K., Fujii Y., Ozaki K., Hirao M., Omori Y., Ota T., Suzuki Y.,
Obayashi M., Nishi T., Shibahara T., Tanaka T., Nakamura Y.,
Isegaki T., Sugano S.;
RA "NEDO human cDNA sequencing project";
RT Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
[5]
RN SEQUENCE OF 1-114 FROM N.A.
RP TISSUE=Blood;
RA Ye M., Zhang O.H., Zhou J., Shen Y., Wu X.Y., Guan Z.Q., Wang L.,
Ren H.Y., Mao Y.F., Dai M., Huang O.H., Chen S.J., Chen Z.;
RT "Human partial CDS from cd34+ stem cells";
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
[6]
RN SEQUENCE OF 437-2426 FROM N.A. (ISOFORM B).
RP TISSUE=Brain;
RC MEDLINE=99397452; PubMed=10470851;
RX Kikuno R., Nagase T., Ishikawa K.-I., Hirose M., Miyajima N.,
Tanaka A., Kotani H., Nomura N., Ohara O.;
RA "Prediction of the coding sequences of unidentified human genes. XIV.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro";
RL DNA Res. 6:197-205(1999).
[7]
RN SEQUENCE OF 554-2426 FROM N.A. (ISOFORM A).
RP MEDLINE=92049296; PubMed=1944255;
RX Chumakov I.M., Berdichevskii F.B., Sokolova N.V., Reznikov M.V.,
Pisolsky V.S.;
RA "Identification of a protein product of a novel human gene SON and
RT its biological effect upon administering a changed form of this gene
RT into mammalian cells";
RL Mol. Biol. (Mosk) 25:731-740(1991).
[8]
RN SEQUENCE OF 709-1079 FROM N.A. (ISOFORM I).
RP TISSUE=Placenta;
RC MEDLINE=93062885; PubMed=1435774;
RX Bliskovskii V.V., Kirillov A.V., Zakhariev V.M., Chumakov I.M.;
RA "The human son gene: the large and small transcripts contains various
RT 5'-terminal sequences";

RL Mol. Biol. (Mosk) 26:807-812(1992).
[9]
RN SEQUENCE OF 1009-1131 FROM N.A.
RP TISSUE=Placenta;
RC MEDLINE=93062884; PubMed=1435773;
RX Bliskovskii V.V., Berdichevskii F.B., Tkachenko A.V., Belova M.E.,
Chumakov I.M.;
RA "Coding part of the son gene small transcript contains four areas of
RT complete tandem repeats";
RL Mol. Biol. (Mosk) 26:793-806(1992).
[10]
RN SEQUENCE OF 1145-2426 FROM N.A. (ISOFORM F).
RP MEDLINE=93048367; PubMed=1424986;
RX Mottioni T., Hume C.R., Konigorski S., Hayes P., Osterweil Z.,
Lee J.S.;
RA "A cDNA clone for a novel nuclear protein with DNA binding
RT activity";
RL Chromosoma 101:618-624(1992).
[11]
RN SEQUENCE OF 1692-2175 FROM N.A. (ISOFORM A).
RP MEDLINE=89039788; PubMed=3054499;
RX Berdichevskii F.B., Chumakov I.M., Kiselev L.L.;
RA "Decoding of the primary structure of the son3 region in human
RT genome: identification of a new protein with unusual structure and
RT homology with DNA-binding proteins";
RL Mol. Biol. (Mosk) 22:794-801(1988).
[12]
RN SEQUENCE OF 1939-2426 FROM N.A. (ISOFORM J).
RP TISSUE=Cerebellum;
RC MEDLINE=99439804; PubMed=10509013;
RX Greenhalf W., Lee J., Chaudhuri B.;
RA "A selection system for human apoptosis inhibitors using yeast";
RT Yeast 15:1307-1321(1999).
[13]
RN SEQUENCE OF 15307-15311(1999).
RT "FUNCTION: Represses hepatitis B virus (HBV) core promoter activity
RT and transcription of HBV genes and production of HBV virions.
RT Binds to the consensus DNA sequence: 5'-GAGTGAAC[AG]CC-3'.
RT Might protect cells from apoptosis. Might be involved in pre-mRNA
RT splicing (by similarity).
CC -1 SUBCELLULAR LOCATION: Nuclear with a speckled distribution.
CC -1 ALTERNATIVE PRODUCTS: 10 isoforms: A, B, C, D, E, F (shown here),
CC G, H, I and J; may be produced by alternative splicing.
CC -1 TISSUE SPECIFICITY: Widely expressed, with the higher expression
CC seen in leukocyte and heart.
CC -1 DOMAIN: Contains 8 types of repeats which are distributed in 3
CC regions.
CC -1 MISCELLANEOUS: Colocalizes with the pre-mRNA splicing factor
CC SFN2/SC-35.
CC -1 SIMILARITY: CONTAINS 1 G-PATCH DOMAIN.
CC -1 SIMILARITY: CONTAINS 1 DBM (DOUBLE-STRANDED RNA-BINDING) DOMAIN.
CC -1 CAUTION: ISOFORM A SEQUENCE FROM REF.7 DIFFERS FROM THAT SHOWN
CC DUE TO A FRAMESHIFT.
CC -1 CAUTION: ISOFORM F SEQUENCE FROM REF.10 DIFFERS FROM THAT SHOWN
CC DUE TO A FRAMESHIFT.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by, and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: AF380179; AL34497.1; -
DR EMBL: X63753; CAA45282.1; ALU_FRAME.
DR EMBL: M36428; AAA3624.1; -
DR EMBL: AF380180; AAL34498.1; -
DR EMBL: AF380181; AAL34499.1; -
DR EMBL: AF380182; AAL34500.1; -
DR EMBL: AF380183; AAL34501.1; -
DR EMBL: AF380184; AAL34502.1; -
DR EMBL: AY026895; AAK07692.1; -
DR EMBL: AF435977; AAL30810.1; -
DR EMBL: X63751; CAC69885.1; -

```

152 AGGAAGACCCCGCGGCATCGTCCCAATATGCGGAGCGGGTTTGAAAC 201
    |||:::||||:::||||::: ||| |||||::: ::
1871 ArgatgserSeratgSeratgSerlySeratrglyArgArgSerAlaE 188
    |||:::||||:::||||::: ||| |||||::: ::
202 CCCGACACGCGAGACGGCGCAAGCCGTTTGGCGAAGCGCAAAATGCGG 251
    |||:::||||:::||||::: ||| |||||::: ::
1887 rlySGlulysArglyArgSerTo..... 1894
    |||:::||||:::||||::: ||| |||||::: ::
252 TTTCGAACTTGCCCGCCCGCTTTTCAAAAACCGAGACATCGAAACAA 301
    |||:::||||:::||||::: ||| |||||::: ::
1896 .....LysHisArgSerlySeratrglu 1903
    |||:::||||:::||||::: ||| |||||::: ::
302 TGTTCAAGCGGTACACGGCTGGGAACAGTCACAGGCTTTGGACAAG 351
    |||:::||||:::||||::: ||| |||||::: ::
1904 ...ArgLyArgLyArgSerSeratrgaBaphanArgLyStbValAr 1919
    |||:::||||:::||||::: ||| |||||::: ::

```

RP SEQUENCE FROM N.A.
RX MEDLINE=94333813; PubMed=7520012;
RT Skehel P.A., Bartsch D.;
RI "Characterization of a Y-Box factor from *Aplysia californica*.";
RL Gene 145:231-235(1994).
CC -1- FUNCTION: BINDS RNA IN VITRO.
CC -1- SUBCELLULAR LOCATION: Nucleolus (Probable).
CC -1- PTM: PHOSPHORYLATION ACTIVATES IN VITRO RNA-BINDING
CC (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE COLD-SHOCK DOMAIN (CSD) FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: U02684; AAA60373.1; -
 DR HSP; P41016; ICGO.
 DR InterPro: IPR002059; Cold_shock.
 DR Pfam: PF00313; CSD; 1.
 DR PRINTS: PR00050; COLDSHOCK.
 DR PRODOM: PD000621; Cold_shock; 1.
 DR SMART: SM00357; CSD; 1.
 DR PROSITE: PS00352; COLD_SHOCK; 1.
 DR Transcription regulation; RNA-binding; Nuclear protein;
 KW Phosphorylation.
 FT MOD_RES 35 99
 FT MOD_RES 226 226 PHOSPHORYLATION (BY CAPK) (POTENTIAL).
 FT DOMAIN 116 138 ARG-RICH REGION (BASIC).
 FT DOMAIN 152 173 ARG-RICH REGION (BASIC).
 FT DOMAIN 186 204 ARG-RICH REGION (BASIC).
 FT DOMAIN 234 253 ARG-RICH REGION (BASIC).
 FT SEQUENCE 253 AA; 29439 MW; AA3400013EAEECE3 CRC64;

alignment_scores:
 Quality: 105.00 Length: 169
 Ratio: 1.265 Gaps: 10
 Percent Similarity: 49.112 Percent Identity: 28.402

alignment_block:
 US-09-303-518D-571 x YBFH_APLCA ..

Align seg 1/1 to: YBFH_APLCA from: 1 to: 253

420 GCACCTTCCTT.....CCACCTGACCGCATGTACAGCGCCGCAAA 463
 96 Alalalsnvalthrglproglnglyserasvalglnlglysetrlsyt 112
 464 TCAAGCGATGACAAAT.....C 483
 112 rlaalalsparargargrpheargarglygltprtyrproargphea 129
 484 ATGACGCGGCGAGGTGCGCGCAAGCAAGCGCCGCGCCGCGCAT 533
 129 rgllygllyarglygltprtyrproargphea 141
 534 ACAAGGCGTCAACCAATCAAGCGCGCGCGCGCGCGCGCAACCA 583
 142 AspsaglialproaspheleProserProarglyargly..... 156
 584 TCATCCTCCGACGACGCTCCCTTCGCGAGGAGCG..... 623
 157Arglyargprotyrtyrglinsargargtyrphelgyp 170
 624CGGCGTGTGGCGGATTTTTCGCGCAACCTGC..... 656
 170 roproargarglygltglt.....Arginsytrleuglglgltglt 184
 657ATACACCATGACACGCGCGCAAAATTTGGCAGCAGTCAAG 697
 185 Glutyrglnleuglnargaspelnglyphearg.....Glyalargargpr 200
 698 GGGTGAACCCCTGT.....T 714
 200 opherlyrargproleuleargthtrSerGlnlyleuleuargargr 217
 715 TTCTGTGCGAAGCGCTGCGCGACGACAGCTTCGTTGCGACATCG 764
 217 rpleuleuargleuargargthtrGlnlyargthtrSerGlnala 233
 765CCCGTCGAAGGGGAATGACGCGCAACCAAGCCACGATG 805
 234 ArgargarglinalrgProtrpelyleuProglinalrglnalrgprolyspr 250
 806 CGGCGCT 812
 250 oargln 252

seq_name: SwissProt_40:CYF_CHLVU

seq_documentation_block:
 ID CYF_CHLVU STANDARD; PRT; 341 AA.
 AC P56316;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Apocytochrome F precursor.
 GN PETA.
 OS Chlorella vulgaris.
 OG Chloroplast.
 OC Eukaryota; Viridiplantae; Chlorophyta; Trebouxiophyceae; Chlorellales;
 OC Chlorellaceae; Chlorella.
 OX NCBI_TaxID:3077;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-IAM C-27 / TAMITA.
 RX MEDLINE-97303241; PubMed-9159184;
 RA Wakasugi T., Nagai T., Kapoor M., Tsudzuki T., Suzuki Y., Hamada A., Ohta T.,
 RA Tsudzuki J., Nakashima K., Tsudzuki M., Sugita M.,
 RA Inamura A., Yoshinaga K., Sugita M.,
 RT "Complete nucleotide sequence of the chloroplast genome from the
 RT green alga Chlorella vulgaris: the existence of genes possibly
 RT involved in chloroplast division".
 RL Proc. Natl. Acad. Sci. U.S.A. 94:5967-5972(1997).
 CC -!- FUNCTION: TRANSLOCATES PROTONS ACROSS THE THYLAKOID MEMBRANE AND
 CC TRANSFERS ELECTRONS FROM THE RIESKE IRON-SULFUR PROTEIN AND PASSES
 CC THEM TO PLASTOCYANIN; THIS FUNCTION IS VERY SIMILAR TO THAT OF
 CC MITOCHONDRIAL CYTOCHROME C1.
 CC -!- SUBUNIT: MEMBER OF THE CYTOCHROME B6/F COMPLEX INCLUDING
 CC CYTOCHROME B6, CYTOCHROME F AND PROBABLY AN IRON SULFUR PROTEIN.
 CC -!- SUBCELLULAR LOCATION: Chloroplast thylakoid membrane (Probable).
 CC -!- SIMILARITY: BELONGS TO THE CYTOCHROME C FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: AB001684; BAA57987.1; -
 DR HSP; P36438; 1HCZ.
 DR InterPro: IPR002325; Apocyl-F.
 DR InterPro: IPR000345; CyfC_heme_bind.
 DR Pfam: PF01333; Apocytochrome_F_1.
 DR PRINTS: PR00610; CYTOCHROME_C_1.
 DR PROSITE: PS00190; CYTOCHROME_C_1.
 DR Electron transport; Heme; Chloroplast; Thylakoid;
 KW Photosynthesis; Photosystem I; Photosystem II; Transit peptide;
 KW Transmembrane.
 FT TRANSIT 1 56
 FT CHAIN 57 341
 FT METAL 57 57
 FT BINDING 77 77
 FT BINDING 80 80
 FT METAL 81 81
 FT TRANSMEM 307 327
 FT SEQUENCE 341 AA; 37350 MW; FBC53FC5D09465F CRC64;

alignment_scores:
 Quality: 105.00 Length: 311
 Ratio: 0.745 Gaps: 13
 Percent Similarity: 45.338 Percent Identity: 21.543

alignment_block:
 US-09-303-518D-571 x CYF_CHLVU ..

Align seg 1/1 to: CYF_CHLVU from: 1 to: 341

```

7  CGTTTAAATTCAGGCTGTTTCCCTTTCGCAACCGCATGCATCATCT 56
   : : : : : : : : : : : : : : : : : : : : : : : : : :
6  LysLeuGluPheAsnPheIleProAsnLeuLys.....LysHisAlaVal 20
   : : : : : : : : : : : : : : : : : : : : : : : : : :
57 GTTGACCGCCCTGCTCAAAATGCTCTCCCTGCTGCTGCTTCTGCTGTC 106
   : : : : : : : : : : : : : : : : : : : : : : : : : :
20 LProSerPheTrpGlyGlnAsnGlnAlaSnIleLeuLysPheSerThrLeu 37
   : : : : : : : : : : : : : : : : : : : : : : : : : :
107 ACACGCTGGGAAACCGGCTCGACATCTGGGCTTTTACTCTTTAAAGGA 156
   : : : : : : : : : : : : : : : : : : : : : : : : : :
37 aLSeLysGlyValLeuValLeuValLysSerPhePheLeuThrAlaSer 53
   : : : : : : : : : : : : : : : : : : : : : : : : : :
157 GACCGCGCGCGCATGTC.....GCCAATATGCGGCA 188
   : : : : : : : : : : : : : : : : : : : : : : : : : :
54 SerAlaAlaLysTrpIlePheAlaGlnGlnAsnTrpAlaAsnProArg 70
   : : : : : : : : : : : : : : : : : : : : : : : : : :
189 GCGG..... 192
   : : : : : : : : : : : : : : : : : : : : : : : : : :
70 uAlaAsnGlyArgIleValCysAlaAsnGlyHisLeuAlaGluLysPro 87
   : : : : : : : : : : : : : : : : : : : : : : : : : :
193 .....GTTTGAACCCCGACAG..... 210
   : : : : : : : : : : : : : : : : : : : : : : : : : :
87 IeGluIleGluValProGlnAlaValLeuProAsnThrValPheGluAla 103
   : : : : : : : : : : : : : : : : : : : : : : : : : :
211 .....CAGACGCTCAAAAGCGTTTTCGCGAAG 239
   : : : : : : : : : : : : : : : : : : : : : : : : : :
104 ValValLysIleProLysPheLysGlnIleLysGlnValLeuAlaSnG 120
   : : : : : : : : : : : : : : : : : : : : : : : : : :
240 GCGAAA.....TGGCGTT 253
   : : : : : : : : : : : : : : : : : : : : : : : : : :
120 LysLysGlyAspLeuAsnValGlyAlaValLeuIleLeuProAsnGly 137
   : : : : : : : : : : : : : : : : : : : : : : : : : :
254 TGGAACTTGGCCCGCGCTTTTCAAAAACCGGAGACATCGAACAATG 303
   : : : : : : : : : : : : : : : : : : : : : : : : : :
137 heGluIleAlaPro.....ProAspArgIleProGluGlu 148
   : : : : : : : : : : : : : : : : : : : : : : : : : :
304 TTCAAAACGGGTACACGGCTGGGAACAGTGCAGACGCTTTGGACAAGG 353
   : : : : : : : : : : : : : : : : : : : : : : : : : :
149 MetLysAlaLysValGlyLysLeuTrpPheGlnProLysSerAlaGlu 165
   : : : : : : : : : : : : : : : : : : : : : : : : : :
354 CGAAGGCTGCTCTTTCATCAGCCCGCACATCGCAGTCAGTATGGGCG 403
   : : : : : : : : : : : : : : : : : : : : : : : : : :
165 sLysThrIlePheValIleGlyProValProGlyLys..... 177
   : : : : : : : : : : : : : : : : : : : : : : : : : :
404 GACGCTACATCAGCCAGCAGCTTCGCTCCACCTGACCGCCATGTACAAG 453
   : : : : : : : : : : : : : : : : : : : : : : : : : :
178 ..LysTrpSerGluMetValPheProIle.....LeuSerProAsp 190
   : : : : : : : : : : : : : : : : : : : : : : : : : :
454 CCGCCGAAATCAAAAGCATAGCAAAATC.....ATGCAGGCGG 494
   : : : : : : : : : : : : : : : : : : : : : : : : : :
191 ProAlaLysThrLysSerIleSerTrpLeuLysTrpProIleTrpAla 207
   : : : : : : : : : : : : : : : : : : : : : : : : : :
495 CAGGCTGCGCGCAAGGCAAAACCGCGCCACCGCATACAGGGGTCA 544
   : : : : : : : : : : : : : : : : : : : : : : : : : :
207 yGlyAsnArgLysArgLysGlnValTrpProAspLysSerLysSerAsn 224
   : : : : : : : : : : : : : : : : : : : : : : : : : :
545 AACCAATATCAAGGCTGCGCGCGCGGCGAGGACATCATCTGCTGCC 594
   : : : : : : : : : : : : : : : : : : : : : : : : : :
224 sThrThrIlePheThrAlaSerAlaIleGlyLysIleThrAlaIle.... 238
   : : : : : : : : : : : : : : : : : : : : : : : : : :
595 GACGAGCTGCTCTCCGCGAGGAAGCGCGCGCTGTGGCGGATTTT 644
   : : : : : : : : : : : : : : : : : : : : : : : : : :
239 .....GluProAlaGlyLysLysGlyGly..... 246
   : : : : : : : : : : : : : : : : : : : : : : : : : :
645 CCGCAACCTCGATACACCATGACACTGGGCGCAAAATGTGCACAGCTCA 694
   : : : : : : : : : : : : : : : : : : : : : : : : : :
247 .....TyrThrLeuThrIle.....GluThrAlaAsnGly 257
   : : : : : : : : : : : : : : : : : : : : : : : : : :
695 AAGCGTGAATAACCTGTTTTCGCTGCGAAGCCCTGCGCGAGCAACA 744
   : : : : : : : : : : : : : : : : : : : : : : : : : :
257 InsThrIleSer.....GluLysLeuProGlyPro 267

```

```

745 GGCTTGCTGTCACATCCGCCGCTCCAGGG 777
   : : : : : : : : : : : : : : : : : : : :
268 GluLeuValValAsnIleGlyAspIleValGly 278

```

seq_name: SwissProt.40:5E5_RAT

seq_documentation_block:

```

ID      5E5_RAT      STANDARD;      PRT;      825 AA.
AC      Q63003;
DT      01-NOV-1997 (Rel. 35, Created)
DT      01-NOV-1997 (Rel. 35, Last sequence update)
DT      01-NOV-1997 (Rel. 35, Last annotation update)
DE      5E5 antigen.
OS      Rattus norvegicus (Rat).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX      NCBI_TaxID=10116;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=WISTAR; TISSUE=Brain;
RX      MEDLINE=96015159; PubMed=8537300;
RA      Suzuki E., Kojima N., Yoshimura K., Uyemura K., Ohta K., Akagawa K.;
RT      "Cloning and sequence analysis of cDNA for a possible DNA-binding
RT      protein 5E5 in the nervous system.";
RL      J. Biochem. 118:122-128(1995).
CC      FUNCTION: MIGHT HAVE DNA-BINDING ABILITY.
CC      SUBCELLULAR LOCATION: Nuclear.
CC      TISSUE SPECIFICITY: EXPRESSED IN NEURONS.
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
DR      EMBL: D37934; BA07153.1; -
KW      DNA-binding; Nuclear protein; Antigen.
SQ      SEQUENCE 825 AA; 86831 MW; AF667FE2FD55BDF CRC64;

```

alignment_scores:

	Quality:	105.00	Length:	309
	Ratio:	0.854	Gaps:	13
	Percent Similarity:	39.806	Percent Identity:	23.625

alignment_block:

US-09-303-518D-571 x 5E5_RAT ..

```

Align seg 1/1 to: 5E5_RAT from: 1 to: 825
6  TCGTTTAAATTCAGGCTGTTTCCCTTTCGCAACCGCATGCATCTCC 55
   : : : : : : : : : : : : : : : : : : : : : : : : : :
535 SerGlnThrLeuProAlaLeuAlaGlyAlaProThrAlaHisAla 551
   : : : : : : : : : : : : : : : : : : : : : : : : : :
56 TGGTACGCGCCCTGCTCAAAATGCTCTCCCTGCTGCTTCTGCTGTC 105
   : : : : : : : : : : : : : : : : : : : : : : : : : :
551 aValProGlyPro..... 555
   : : : : : : : : : : : : : : : : : : : : : : : : : :
106 CACACGCTGGGAAACCGGCTGGACATCTGCGCTTTTACCTTTAAAGGA 155
   : : : : : : : : : : : : : : : : : : : : : : : : : :
556 .....GlyProAlaAlaIleThrLeuGlyLysArgLysArgGly 569
   : : : : : : : : : : : : : : : : : : : : : : : : : :
156 AGACGCGCGCGCATCTCGCCCAATATGCGGCGGCGGTTTAAACCCG 205
   : : : : : : : : : : : : : : : : : : : : : : : : : :
570 sThrTrpArgLysArgArgGlyGlyAlaGly..... 581
   : : : : : : : : : : : : : : : : : : : : : : : : : :
206 ACAGCGTCAAGGCTCAAAAGCGTTTTCGCGAAGCGCAAAATCGGTTTG 255
   : : : : : : : : : : : : : : : : : : : : : : : : : :
582 ...AlaSerLysLysGlyArg...GlyGlyArgGlyArgGlyArg.... 594
   : : : : : : : : : : : : : : : : : : : : : : : : : :
256 GAACCTGCGCCCGGTTTTCAAAACCGGAAAGACATGCAACAATGTT 305

```


[illegible]

seq_name: SwissProt_40:YT44_STRFR

```

seq_documentation_block:
ID      YT44_STRFR      STANDARD;      PRT;      395 AA.
AC      P20188;
DT      01-FEB-1991 (Rel. 17, Created)
DT      01-FEB-1991 (Rel. 17, Last sequence update)
DT      16-OCT-2001 (Rel. 40, Last annotation update)
DE      Hypothetical 44.4 kDa protein in transposon Tn4556.
OS      Streptomyces fradiae.
OC      Bacteria; Firmicutes; Actinobacteria; Actinobacteriidae;
OC      Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
ON      NCBI_TaxID=1906;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      TRANSPOSON-Tn4556;
RX      MEDLINE=90185236; Pubmed=2155856;
RA      Slemieniak D.R., Slightom J.L., Chung S.T.;
RT      Tn4556, a class-II transposon related to Tn3;
RL      Gene 86:1-9(1990).
-----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
EMBL; M29297; AAA88562.1; -
PTR; J00430; J00430.

```

KW Hypothetical protein; Transposable element.
SQ SEQUENCE 395 AA; 44379 MW; C50B556F0E3B2838 CRC64;

```

alignment_scores:
    Quality: 104.50
    Ratio: 1.025
    Percent Similarity: 37.363
    Length: 273
    Gaps: 14
    Percent Identity: 23.443

```

```
alignment_block;
US-09-303-518D-571 x YT44_STRFR ..
```

Align seg 1/1 to: YT44_STRFR from: 1 to: 395

[illegible]

705 AACCTGTTTCTGCTGCGACCCCTCCGACGACGACGCTGCTG 754
 240LeuLeuArgProLeuGlnArgProLeuHisLeuArgIn 252
 755 TGCACATCCGCCCGCTCCA 773
 252 rArgHisProHisArgPro 258
 seq_name: SwissProt_40:VE2_HPV05

seq_documentation_block:
 ID VE2_HPV05 STANDARD; PRT; 514 AA.
 AC P06921;
 DT 01-JAN-1988 (Rel. 06, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE Regulatory protein E2.
 GN E2
 OS Human papillomavirus type 5.
 OC Viruses; dsDNA viruses, no RNA stage; Papillomaviridae;
 OC Papillomavirus.
 OX NCBI_TaxID=10578;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=87207670; PubMed=3033892;
 RA Zachow K.R., Ostrow R.S., Faras A.J.;
 RT Nucleotide sequence and genome organization of human papillomavirus
 type 5.
 RL Virology 158:251-254 (1987).
 CC -1- FUNCTION: E2 REGULATES VIRAL TRANSCRIPTION AND DNA REPLICATION.
 CC IT BINDS TO THE E2RE RESPONSE ELEMENT (5'-ACCGNNNNCGT-3') PRESENT
 CC IN MULTIPLE COPIES IN THE REGULATORY REGION. IT CAN EITHER
 CC ACTIVATE OR REPRESS TRANSCRIPTION DEPENDING OF E2RE'S POSITION
 CC WITH REGARDS TO PROXIMAL PROMOTER ELEMENTS. REPRESSION OCCURS
 CC BY STERICALLY HINDERING THE ASSEMBLY OF THE TRANSCRIPTION
 CC INITIATION COMPLEX. THE E1-E2 COMPLEX BINDS TO THE ORIGIN OF DNA
 CC REPLICATION.
 CC -1- SUBUNIT: BINDS DNA AS A DIMER.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC DR EMBL: M17463; AAA46986.1; -
 CC DR PIR: D26277; W2ML5.
 CC DR HSP: P03122; ZBOP.
 CC DR InterPro: IPR001866; E2_N.
 CC DR InterPro: IPR000427; Early2_C.
 CC DR Pfam: PF00511; E2_C; 1.
 CC DR Pfam: PF00508; E2_N; 1.
 CC DR ProDom: PD000672; Early2_C; 1.
 CC DR ProDom: PD000678; E2_N; 1.
 CC DR Early protein, transcription regulation; Activator; DNA-binding;
 CC Trans-acting factor; DNA replication; Repressor; Nuclear protein;
 CC KW SEQUENCE 514 AA; 57532 MW; 0100722143BF7256 CRC64;
 CC SQ

alignment_scores:
 Quality: 104.50 Length: 319
 Ratio: 0.711 Gaps: 17
 Percent Similarity: 46.082 Percent Identity: 23.197

alignment_block:
 US-09-303-518D-571 x VE2_HPV05 ..
 Align seg 1/1 to: VE2_HPV05 from: 1 to: 514

44 CCATGCACATCTGTTGACCCCGCTGCTCA..... 73
 197 ProValThrSerSerThrProProGlySerProGlyGlyAlaAspTh 213
 74 .AANGCCTTCGCCCTGCTGCTTCCTGCTGCACACGCTGGAAACG 122
 213 rAsnThrThrProAlaThrProThrThrSerThrThrAlaValAspSer 230
 123 GCTGGACATCTGGCGTTTACCTTTAAGGAAGACCGCGCGCATCG 172
 230 hrSer.....ArgGlnLeuThrThrSerLys 238
 173 TCGCCATATATGCGAGCGGGTTTGAACCCGACGACGACGCTCAAA 222
 239 GlnPro.....GlnGlnThrGlnThrArgGlyLar 248
 223 GCGGTTTTTGGGAACGCGCAAAATGCGGTTTGAACCTTGCCTCCGCGGT 272
 248 gArgTyrGlyArgArgProSer..... 255
 273 TTTCAAAAACCGGAAGACATCGAACAATGTTCAACGCGTACGCGCT 322
 256 ..SerLysSerArgArgSerGlnThrGlnArgArgSerArgSerArg 271
 323 GGGAAACGCTGCAGCAGCGCTTGGACAAAGGCGGAAGGCTGCTTCATC 372
 272 HisArgSerArgSerArgSerArgSerArgSerLys.....SerG 285
 373 AGCGCCGACATGCGCAGCATTTGGCGGACGCTACATCAGCCAGCA 422
 285 nThrHisThrThrArgSerThrThrArgSerArgSerThrSerLeuThr 302
 423 GCTTCGTTCCACCTGACCGCCATGTACAGCCCGCAAAATCAAG... 469
 302 yThrThrArgAla.....LeuThrSerArgSerArgSerArgGly 314
 470 CGATAGCAAAATCATCAGGCGGCGAGGCTGCGCGG..... 506
 315 ArgSerProThrThrArgSerArgGly.GlyGlyArgSerProArgArg 331
 507CAAGCGCAAAAC 518
 331 rGSerArgSerProSerThrThrSerSerCysThrThrThrInArgSerGln 347
 519 CGCGCCGAC.....CGGCATACAGGCGTCAAAACAA 550
 348 ArgAlaArgAlaGluSerSerThrThrArgGlyAlaArgGlySerArg 364
 551 TCATCAAGGCGCTGCGCGC...GGCGAGGCAACATCAT..... 587
 364 ySerArgGlyLysSerArgGlyGlyArgArgArgArgArgArgSer 381
 588CGTCCGACACGCTGCTTCGCGAGGAAGCGCG 623
 381 erSerSerSerProAlaHisLysArgSerArgGlyGlySerAlaLys 397
 624 CGCGCT.....GTGGCGGATTTTTCGCAACCTG 655
 398 LeuArgGlyValSerProGlyGluValGlyLysSerLeuArgSerVal 414
 656 CATACCATCATGACTGCGCGCAAAATTTGGC.....ACAGTCAAGGC 699
 414 rSerLysHisThr.....GlyArgLeuGlyArgLeuLeuGlnGluAla 429
 700 GTGAACACCTGTTTCTGCTGCGAAGCGCTGCGCA..... 737
 429 rGAspProProValIleIleValLysGlyAlaIleAsnThrLeuLysAsn 445
 738 ...CGACAGGCTTCTGTTGCACATCCGCCCGCTCAAG..... 776
 446 ValArgAsnArgAlaLysIleLysTyrMetGlyLeuPheArgSerPhe 462
 777GGAATTGAACGCGCAACAAAGCC 798

Percent Similarity: 49.751 Percent Identity: 25.373

alignment_block:

US-09-303-518D-571 x SON_MOUSE ..

Align seg 1/1 to: SON_MOUSE from: 1 to: 2404

```

230 TTGGCGAAGCGCAAAATCGGTTGGAACTTGGCCCGCGTTTTCGAA 279
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:
1807 LeuArgSerArgSerArgSerArgSerArgSerArgSerArgSerArg 1823
280 AAACCGGAGACATCGAACAATGTTCAAGCGGTACAGCGCGGAGACA 329
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:
1823 SArgThrSerGluSerArgSerArgAlaArgLysArgSerSerL 1840
330 CGTGACGAGCGTTTGGACAGAGCGGCGGCTGCTTTCATCCGCGCG 379
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:
1840 YSerHisArgSerGlnThrArgSerArgSerArgSerArgArgArg 1856
380 ACATCGGACGCTACGATTGGCGGACGCTACATCAGCCAGCAGCTTCCG 429
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:
1857 ArgSer.....SerArgSerArgSerly 1864
430 TTCACCTGACCGCCATGACAGCCCGCAAAATCAAGCATGACACA 479
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:
1864 SSerArgGlyArgArgSerArgSerArgSerArgSerArgSerArg 1881
480 AATCATCAGGCGGCGGAGGTGCGCGCAAAAGCAACCGCGCCACCG 529
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:
1881 YS..HisArgSerLysSerArgGluArgLysArgLysArgSerSerAr 1897
530 GCATACAGG.....GTCACAAACAATCATCAGGCGCTGCGCGCGGC 573
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:
1897 GAspAsnArgLysAlaAlaArgAlaArgSerArgThrProSerArg 1914
574 GAGGACACCATCATCTGCC...CGACACGCTCCCTTCGCGAGAGAG 620
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:
1914 eArgSerHisThrProSerArgArgArgArgSerLysSerValGlyArg 1930
621 CGGCGCGCTGTGGCGGATTTTTCGCAAACTGCTACATCAGTACAC 670
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:
1931 ArgArgSerPhe.....ArgSerArgThrProSerArgArgArg 1934
671 TGGCGGCAAAATGTCACACGTCGCAAGCGTGAACCTGTTTTCCTGC 720
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:
1949 eArgArgThrProSerArgArgSerArgThrProSerArgArgSerArg 1965
721 TGGCAACGCTGCGCGACGACAGAGCTT.....CGTGTTCGACAT 761
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:
1949 eArgArgThrProSerArgArgSerArgThrProSerArgArgSerArg 1965
762 CCGCCCCCTCCCAAGGGAATTGAAGCGGACCAACCCAGATGCCCGC 811
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:
1966 ProSerArg.....ArgSerArgThrProSerArgArgArg 1977
812 T 812
1977 g 1977
seq_name: SwissProt_40:SFR2_CHICK
seq_documentation_block:
ID SFR2_CHICK STANDARD: PRT; 221 AA.
AC P30352:
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Splicing factor, arginine/serine-rich 2 (Splicing factor SC35) (SC-35)
DE (Splicing component, 35 kDa) (PR264 protein).
GN SFRS2.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;

```

```

OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Thymus;
RX MEDLINE=92212859; PubMed=1557353;
RA Vellard M., Sureau A., Soret J., Martinerie C., Perbal B.;
RT "A potential splicing factor is encoded by the opposite strand of the
RT trans-spliced c-myc exon.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:2511-2515(1992).
CC -1- FUNCTION: NECESSARY FOR THE SPLICING OF PRE-MRNA. IT IS REQUIRED
CC FOR FORMATION OF THE EARLIEST ATP-DEPENDENT SPLICING COMPLEX AND
CC INTERACTS WITH SPLICOSOMAL COMPONENTS BOUND TO BOTH THE 5' AND 3'
CC SPLICING SITES DURING SPLICOSOME ASSEMBLY. IT ALSO IS REQUIRED FOR
CC ATP-DEPENDENT INTERACTIONS OF BOTH U1 AND U2 SNRNPs WITH PRE-
CC mRNA.
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- ALTERNATIVE PRODUCTS: DIFFERENT FORMS OF THE PROTEIN MAY BE
CC PRODUCED BY ALTERNATIVE SPLICING.
CC -1- PTM: EXTENSIVELY PHOSPHORYLATED ON SERINE RESIDUES IN THE RS
CC DOMAIN (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE SR FAMILY OF SPLICING FACTORS.
CC -1- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RRM).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC of the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X62446; CAA4306.1; .
DR PIR: S17327; S17327.
DR PIR: B42701; B42701.
DR HSP: P09651; 1UP1.
DR InterPro: IPR00504; RRM.
DR Pfam: PF00076; rrm; 1.
DR SMART: SM00360; RRM; 1.
DR PROSITE: PS50102; RRM; 1.
DR PROSITE: PS00030; RRM_RNP_1; 1.
DR Nucleic acid protein: RNA-binding; mRNA splicing; Alternative splicing;
KW Phosphorylation.
FT DOMAIN 14 92 RNA-BINDING (RRM).
FT DOMAIN 117 221 GLY-RICH (HINSE REGION).
FT DOMAIN 117 221 ARG/SER-RICH (RS DOMAIN).
SQ SEQUENCE 221 AA; 25524 MW; 75A4DBF9170F1BF CRC64;

```

alignment_scores: Quality: 102.50 Length: 211
Ratio: 0.995 Gaps: 8
Percent Similarity: 48.815 Percent Identity: 22.749

alignment_block:

US-09-303-518D-571 x SFR2_CHICK ..

Align seg 1/1 to: SFR2_CHICK from: 1 to: 221

```

238 ACTTGGCCCCCGGTTTTCAGAAAACGGAAGACATCGAACAATGTTCA 307
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:
29 ThrLeuArgArgValPheGluLysTyrGlyArgValGlyLysPheVal 45
308 AAGCGGTACAGCGCTGGGAGACAGTGCAGCGCTTTGGA..... 347
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:
45 eProArgAspArgTyrThrLysGluSerArgGlyPheAlaPheValArg 62
347 ..... 347
62 HeHisAspLysArgAspAlaGluAspAlaMetAspAlaMetAspGlyAla 78
348 .....CAAGGCGAAGGCGCTGCTTCATCAGCGCGGACATCGGAGCT 391
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

79 ValLeuaspGlyArgGluLeuArgValGlnMetaLala..... 90
392 ACAGTTTGGCGGAGCTAGATCACCAGCAAGCTTCCCTCCACCTGACC 441
   |||:|||||
91 ArgTyrGlyArg.....ProProsp 98
442 GCCATGTACAGCCCGCAAAATCMAAGCATAGCAAAATCAGCAAGC 491
   |||:|||||
98 eRHs...HisSerArgTyrGlyProProProAlaArgTyrGlySer 113
492 GGGCAGGGGGGGCGGCAAGCAAGCAAGCCCGCCAGCATACAGAGG 541
   |||:|||||
114 GtYtYrGlyArgTyrSerArgSerProArgTyrArgTyrArgTyrSe 130
542 TCMAACAATCATCAAGCCCTCGCGCGCGGCGAGCAACCATCATTCT 591
   |||:|||||
130 ArgSerArgSerArgSerArgSerArgSerArgTyrTyrSerArg 147
592 CCCGACGACGTCCTTCTCCGAGAGAGCGCGCGCTGGGCGGATTT 641
   |||:|||||
147 eRlySerArgSerArgTyrArgTyrArgTyrArg..... 158
642 TTTGGCGCAACCTGCATACACATGACACCTGCGCAAAATTGGCAG 691
   |||:|||||
159 .....SerThSerLysSerArgSerAlaAr 167
692 TCMAAGCGTGAACCCCTTTTCTGCTGCGAGCAGCCTGCCCA...C 738
   |||:|||||
167 gArgSerLysSerLysSerSerValSerArgSerArgSerArg 184
739 GGAACAGCTTCTGTGTGCACATCCGCGCGCGCAAGGGGAATTGAC 788
   |||:|||||
184 rgserArgSerArgSerArgSerProPro...ProThSer...LysArg 198
789 CACAAGAGCCGAGATGCCCGCTGTTCACCG 821
   |||:|||||
199 GluSerArgSerArgSerArgSerPro 209

seq_name: SwissProt_40:FXR2_HUMAN

seq_documentation_block:
ID FXR2_HUMAN STANDARD; PRT; 673 AA.
AC P51116;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Fragile X mental retardation syndrome related protein 2.
GN FXR2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=96080171; PubMed=7489725;
RA Zhang Y., O'Connor J.P., Stomi M.C., Srinivasan S., Dutra A.,
RA Nusbaum R.L., Dreyfuss G.;
RT "The fragile X mental retardation syndrome protein interacts with
RT novel homologs FXR1 and FXR2."
RL EMBO J. 14:5358-5366(1995).
RN [2]
RP SEQUENCE OF 1-27 FROM N.A.
RX MEDLINE=98100442; PubMed=9437788;
RA Joseph D.R.;
RT "The rat androgen-binding protein (ABP/SHBG) gene contains triplet
RT repeats similar to unstable triplets; evidence that the ABP/SHBG and
RL the fragile X-related 2 genes overlap."
CC -1- FUNCTION: RNA-BINDING PROTEIN. INTERACTS WITH FXR1 AND FXR2.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: CONTAINS 2 KH DOMAINS.
CC -1- SIMILARITY: BELONGS TO THE FXR1 FAMILY.

```

```

CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL: U15101; AAC50292.1;
DR EMBL: AF044263; AAC03357.1;
DR HSPB: Q06787; 2FMR.
DR MIM: 605339;
DR InterPro: IPR004087; KH.
DR InterPro: IPR004088; KH_TYPE_1.
DR Pfam: PF00013; KH-domain; 2.
DR SMART: SM00322; KH; 2.
DR PROSITE: PS50084; KH_TYPE_1; 2.
DR RNA-binding; Repeat.
FT DOMAIN 232 261 KH 1.
FT DOMAIN 295 324 KH 2.
FT DOMAIN 414 418 POLY-SER.
FT DOMAIN 544 552 POLY-ARG.
FT DOMAIN 584 594 POLY-ARG.
SQ SEQUENCE 673 AA; 74128 MW; D20FBD8634D0B CRC64;

alignment_scores:
Quality: 102.50 Length: 223
Ratio: 0.967 Gaps: 10
Percent Similarity: 47.534 Percent Identity: 23.767

alignment_block:
US-09-303-518d-571 x FXR2_HUMAN ..
Align seg 1/1 to: FXR2_HUMAN from: 1 to: 673

192 GGGTTGAACCCCGACGACGACGCTCAAAAGCGTTTTCGGGAACG 241
   |||:|||||
391 GlyPheArgProPro.....GlySerG 398
242 CAAATGCGTTGGAACTGCGCCCGGCTTTTCAAAAACCGGAAGAC 291
   |||:|||||
398 YArgGlySerGlyGlySer.....AspYsAlaGlyTyrS 410
292 ATCGAAACAAATGTTTAAAGCCGTACAGCGCTGGAGACAGTCGACAGC 341
   |||:|||||
410 eRThAspGlySerSerSerSerSerLeuHisAlaThrArgThrTyrGly 426
342 TTTGACAGAGGCGAGAGGGCTGTTCATCATCAGCCCGCACATGGCAGCT 391
   |||:|||||
427 GlySerTyrGlyGlyArgGly.....ArgGlyAr 436
392 AGCATTTGGCGGAGCGCTACATCAGCCAGCAGCTTCGCTCCACTGACC 441
   |||:|||||
436 gArgThrGlyGlyProAlaTyrGlyProSerSerAspValSer..... 450
442 GCCATGTACAGCCCGCAAAATCMAAGCATAGCAAAATCATCGAGGC 491
   |||:|||||
451 .....ThrAlaSerGluThrGluSerGluYs..... 459
492 GGGCAGGGTGGCGGCAAGCAAGCAAGCCGCGC.....CA 526
   |||:|||||
460 .....ArgGluGluProAsnArgAlaGlyProGlyAspArgAs 472
527 CCGCATACAGAGGGTCAAAATCATCAAGCCCTGCGCGGCGAG 576
   |||:|||||
472 ProProThrArgGlyGluGluSerArgTyrArgProThrGlyGlyArg 489
577 GCACATCATCTCTGCGCGACGATCCCTTCCGAGAGAGGGGGGG 626
   |||:|||||
489 LysArgGlyProProAlaProArgProThrSerArgTyrAsnSer 505

```

627 CGTGGGCGGATTTTTCGCAACCTGCAACACGACACTGGCG 676
 506 SerIleSerValIleuLYsAspPro..... 514
 677 CAAATTTGGCACAGCTCAAGAGCGTGAACCCGTTTCTCTCGGAA 726
 515AspSerAsnProTyrSerIleuLeuAsp 524
 727 CG.....CTGCCCAGCAGACAGCTTGTGTCATCGCCCGCT 770
 524 hSeSerGluProGluProValAspSerGluProGluProPro... 539
 771 CCAAGGCGAATTGACGCGCAACAAAGCCGCGCTGTTCAC 820
 540 ProAlaSerAlaArgArgArgSerArgArg...ArgArgThrAspG 555
 821 GCATACCGAATATTGAT 839
 555 uAspArgThrValMetAsp 561
 seq name: SwissProt_40:VE2_HP36

seq_documentation_block:
 ID VE2_HP36 STANDARD; PRT; 509 AA.
 AC P50809;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE Regulatory protein E2.
 GN E2.
 OS Human papillomavirus type 36.
 OC Viruses; dsDNA viruses, no RNA stage: Papillomaviridae;
 OC Papillomavirus.
 OX NCBI_TaxID=37957;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA deJuss H.;
 RL Submitted (OCT-1995) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: E2 REGULATES VIRAL TRANSCRIPTION AND DNA REPLICATION.
 CC IT BINDS TO THE E2RE RESPONSE ELEMENT (5'-ACNNNNNGT-3') PRESENT
 CC IN MULTIPLE COPIES IN THE REGULATORY REGION. IT CAN EITHER
 CC ACTIVATE OR REPRESS TRANSCRIPTION DEPENDENT OF E2RE'S POSITION
 CC WITH REGARDS TO PROXIMAL PROMOTER ELEMENTS. REPRESSION OCCURS
 CC BY SPERICALLY HINDERING THE ASSEMBLY OF THE TRANSCRIPTION
 CC INITIATION COMPLEX. THE E1-E2 COMPLEX BINDS TO THE ORIGIN OF DNA
 CC REPLICATION.
 CC -1- SUBUNIT: BINDS DNA AS A DIMER.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: U31785; AAA79439.1; ..
 DR HSSP; P03122; 280P
 DR InterPro; IPR001866; E2_N
 DR InterPro; IPR000427; Early2_C
 DR Pfam; PF00511; E2_C; 1.
 DR Pfam; PF00508; E2_N; 1.
 DR ProDom; PD000672; Early2_C; 1.
 DR ProDom; PD000678; E2_N; 1.
 DR Early protein: Transcription regulation; Activator; DNA-binding;
 KM Trans-acting factor; DNA replication; Repressor; Nuclear protein.
 KW SEQUENCE 509 AA; 57186 MW; BD41FA299E150D91 CRC64;
 SO

alignment_scores:
 Quality: 102.00 Length: 241
 Ratio: 0.836 Gaps: 10

Percent Similarity: 50.622 Percent Identity: 22.822

Alignment_block:
 US-09-303-518D-571 x VE2_HP36 ..

Align seg 1/1 to: VE2_HP36 from: 1 to: 509

185 GGCAGGCGGTTTGAACCCGACGACGCGGTCAACGCGTTTTCG 234
 250 GLyArgArg.....ProSerArgThrArgArgProAlaIal 263
 225 GAACGCGCAAAATCGGTTTGAACCTTGCCCGCGCTTTTCAAAAAC 284
 263 SGLnArgArgSerArg.....SerArgHisArgSerSer 275
 285 GGAACATCGCAACATGTTTCAAGCGGTACACG..... 319
 275 rSerArgSerArgSerArgInserArgSerHisThrProThrArgSer 291
 320GCTGGACACGTCGACGAC 339
 292 AlaThrThrArgSerArgSerProSerLeuAlaIalSerGlyAla 308
 340 GCTTTGGACAGCGCAAGGCTGCTGTATCAAGCGCGCATCGGACG 389
 308 gValSerThrArgSerArgSerArgSerThrSerArgArgGlyAla 325
 390 CTACGATTTGGCGGACGCTACATCAGCAGACGTTCCGTTCCACCTGA 439
 325 rArgArgSerArgSerProSerThrSerSerThrThrThrAs 341
 440 CCGCATGTACAAACCCGCAAAATCAACGATGACAAATCATCGAC 489
 341 nLyArgSerArgValAlaGlyAlaThrThrThrThrThrThrThr 358
 490 GCGGCGAGGTCGCGCGCAAGCAAAACCGCGCCACCGCATCAACAG 539
 358 lYelArgGlyAlaIalArgGlySerGlyGlyAlaArgArgArg 374
 540 GGTCAACAATATCATCAAGCGCTGCGCGCGGCGGACGACATCATC 589
 375 SerSerSerSer...ThrSerProAlaHisLysArgSerArgGluHis 390
 590 TGCCGACACACCTCTCTCCGACGAAAGCGCGCGCTGCGCGCAT 639
 390 rValArgSerArgGlyAlaSerProAsp.....GlnValGlyLys 404
 640 TTTTTCGCAACCTGCATACACCATGACACCTGCGCAAAATTGCG... 686
 404 eLeuArgSerValSerSerLysHisThr.....GlyArgLeuGlyArg 418
 687 ...ACAGCTCAAAAGCGGTGAACCCGTTTTCGTCGCGCAACGCTGC 733
 419 LeuLeuGluGluAlaLeuAspProValIleLeuValArgGlyAla 435
 734 CGGA.....CGGACAAAGCGTTCGTTGTCATCCGCCGCC 768
 435 aAsnThrLeuLysCysPheArgAsnArgAlaLysIleLysTyrMetGly 452
 769 GTCGCAAG.....GGAATT 782
 452 euTyrArgSerPheSerThrThrTrpSerTrpValAlaIalAspGlyThr 468
 783 GAACGCAACAAAGCCACGA 803
 469 GluArgLeuGlyArgProArg 475
 seq_name: SwissProt_40:YH1_EBV
 seq_documentation_block:
 ID YH1_EBV STANDARD; PRT; 660 AA.
 AC P03181;
 DT 21-JUL-1986 (Rel. 01, Created)

21-JUL-1986 (rel. 01, Last sequence update)
 15-JUL-1998 (rel. 36, Last annotation update)
 DE Hypothetical BHLF1 protein.
 OS Epstein-Barr virus (strain B95-8) (Human herpesvirus 4).
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 OC Gammaherpesvirinae; Lymphocryptovirus.
 NCBI_TaxID=103777;
 RN [1]

SEQUENCE FROM N.A.
 MEDLINE=84270667; PubMed=6087149;
 RA Baer R., Bankier A.T., Biggin M.D., Deininger P.L., Farrell P.J.,
 RA Gibson T.J., Hatfull G., Hudson G.S., Satchwell S.C., Seguin C.,
 RA Tuffnell P.S., Barrell B.G.,
 RT "DNA sequence and expression of the B95-8 Epstein-Barr virus genome."
 RL Nature 310:207-211(1984).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; V01555; NOT_ANNOTATED_CDS.
 DR PIR; A03742; O0B83.
 KW Hypothetical protein; Early protein; Repeat.
 FT DOMAIN 148 648 4 X 125 AA TANDEM REPEATS.
 FT REPEAT 149 273 1.
 FT REPEAT 274 398 2.
 FT REPEAT 399 523 3.
 FT REPEAT 524 648 4.
 SO SEQUENCE 660 AA; 66244 MW; 86DA1D67A37152A2 CRC64;

Alignment_scores:

Quality: 102.00 Length: 306
 Ratio: 0.779 Gaps: 13
 Percent Similarity: 42.810 Percent Identity: 25.163

alignment_block:

US-09-303-518D-571 x YHLL_EBV

Align seg 1/1 to: YHLL_EBV from: 1 to: 660

```

27 TCCCTCTTGGCAACCCGACATCTGTGACCGCTGCT..... 71
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
211 ThrProHisProGluArgGlySerGlyProAlaAspProAlaAlaAl 227
72 .....
227 aATgLeuProProGluArgGlnGluProArgLeuProGlnAspLeuAla 244
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
94 CTTTCCTGTGTCACACGCTGGGAACCGGCTCGACATCTGGGTTTAA 143
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
244 lAlaGlnArgGlyProAlaGlyProProProThrArgSerGlyAla... 259
144 CTTTAAAGAGAGACCGCGCGCATGTGTCGAATATGCGGACGCGG 193
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
260 .....AlaAlaGlnArgGlnArgGlnArgGlnArgGlnArgGln 271
194 GTTTAAACCCGACACGACGAGCGGTCAAGACGCTTTGGGGAACGCA 243
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
272 .....ProArgSerAla..... 275
244 AAATGCGGTGGAACTTGGCCCGGCTTTTCAAAAACCGGAGACAT 293
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
276 .....ArgAsnProGlyCysProArg.....ThrTrpArgGly 286
294 CGAAACATGTTCAA..... 308
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
287 ArgSerGlyAlaGlnArgGlnArgGlnArgGlnArgGlnArgGln 303

```

seq_name: SwissProt_40:SGS4_DROME

seq_documentation_block:

ID SGS4_DROME STANDARD; PRT; 297 AA.
 AC 000725; OSN472;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Salivary glue protein Sgs-4 precursor.
 GN Sgs4 OR Sgs-4 OR EG:96G10.6 OR CG12181.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 NX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=OREGON-R, KARGNAS, AND SAMARKAND;
 RX MEDLINE=92223113; PubMed=1562607;
 RA Furia M., Digilio F.A., Ariano D., Favia G., Polito L.C.;
 RT Molecular characterization of a Drosophila melanogaster variant
 RT strain defective in the Sgs-4 gene dosage compensation."
 RL Biochim. Biophys. Acta 1130:314-316(1992).
 SO SEQUENCE FROM N.A.

```

309 .ACGCGTACACGGCTGGGAACACGTGCACACGCTTTGGACAAAGCGGCA 357
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
303 oSerGlyProThrGlyGlyArgProAlaAlaProGlyAlaProGlyThrP 320
358 GGGCTCTCTTCAATCAGCCGCACATCGGACGCTTCGATTTGGGGCGGAC 407
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
320 roAlaAlaProGlyProGlyGlyGlyAlaAlaAlaAlaProSerGlyAlaThr 336
408 CTACATCACCACGACAGCTTCGTTCCACCTGACCGCATGTACAGGCC.. 455
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
337 ProHisProGluArgGlySerGlyProAlaAspProProAlaAlaAlaAr 353
456 .....GCCAAATCAAGGATAGA.....CAAATCATCAGCGC 492
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
353 gleuProProGluArgGlnGluProArgLeuProGlnAspLeuAlaAla 370
493 GGCAGGG.TGGCGGCAAAAGCAAAACCGCGCCCGCATCAAGAGGG 541
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
370 lAlaArgCysProAlaGlyProProProProThrArgSerGlyAlaAlaAla 386
542 TCAAACAAATCATCAGGCGC.....TGGCGCGGCGGCGAGCAACCATC 585
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
387 GlnArgThrHisArgArgProProGlyCysProArgSerAlaArgAsnPr 403
586 ATCCGCGCCGACACGCTCTTCGCGCAGGAAGCGCGCGCTGTGGC 635
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
403 oGlyCysProArgThr...TrpArgArgSerGlyAlaGlnArgGly. 418
636 GGATTTTTCGCAACCTGATACACCATGACACATGGCGCAAAATTTGG 685
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
419 .....HisProProProGlyAlaGlyGlnArgPro 428
686 CACACGTCAAAGCGGTGAACCCGTTTTCGCGCAGCAAGCGCGCC 735
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
429 SerGlyProThrGlyArgProAlaAlaProGlyAlaProGlyThrP 445
736 GACGACAAAGCT.....TCGTGTCACAT 761
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
445 oAlaAlaProGlyProGlyGlyAlaAlaAlaAlaProSerGlyAlaThrP 462
762 CGCGCCGTCGAAGGGAATTGACGCAACAGCCGACGATGGCGCGC 811
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
462 roHisProGluArgGly.....SerGlyProAlaAspProPro 474
812 TGTTCACCGCAATACCG 829
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
475 AlaAlaAlaArgLeuPro 480

```



```

151 gthrglupProthrcysLysThrGluProProthrcysLysThrGluPro 168
264 CCGCGGTTTTCACAAACCGAGACATGCAACAAATGTTCAAGCGG 313
168 roProthrcysLysThrGluProProCysGluLysThrGluLysArg 184
314 TCACGCGGTGGACACGTCGACGAGCTTGGACAGCGGAGG 358
185 IleLysArgHisArgThr.....LysArgThrLysArgSerLysSerTh 199
359 .....G 359
199 rLysLysIleValHisHisHisAsnArgProGlyThrThrProGluSerG 216
360 GCTGC...TGTTTCATCAGCGCCGACATCG.....GAGCTACG 394
216 LysCysGlyCysGlySerLysAsnGluSerGlyGlyGlySerGlyCys 232
395 ATTTGGGCGGCGCTACATCAGCCAGCAGCTTCCACCTGACCGCC 444
233 IleLeuLysAspLeuLeuThr.....Proly 241
445 ATGTACAGCGCCGCAAAATCAAGCGATGACAAATCATGACG... 490
241 scysProAspSerLysProLysProGlnAlaSerProLysCysLysSerA 258
491 .....CGGCGAGGCTGCGCGGCAAGCAAGCAAGCGCCGACCG 529
258 sPrProLysProLysAlaAlaSerLysThrSerLysProLysProLys 274
530 GCATCAAGGCGGTCAACAAATCAAGCGCCCTGCGCGGCGGCAAGCA 579
275 AlacysAspSerGlyLysLysAsnThrThrLysLysProAlaGlySerH 291
580 ACCA 583
291 nPro 292

seq.name: SwissProt_40:MANA_RHOMR
seq_documentation_block:
ID MANA_RHOMR STANDARD; PRT; 1021 AA.
AC P49425.
DT 01-FEB-1996 (Rel. 33, Created)
DT 13-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Mannan endo-1,4-beta-mannosidase (EC 3.2.1.78).
GN MANA.
OS Rhodothermus marinus (Rhodothermus obamensis).
OC Bacteria; CFM group; Rhodothermus group; Rhodothermus.
OX NCBI_TaxID=29549;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 43812;
RA Pollitz O., Krah M., Bortless R.;
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
CC -! CATALYTIC ACTIVITY: Random hydrolysis of 1,4-beta-D-mannosidic
CC linkages in mannan, galactomannans, glucomannans, and
CC galactoglucomannans.
CC -! SIMILARITY: BELONGS TO FAMILY 26 OF GLYCOSYL HYDROLASES.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X90947; CA62442.1; -
CC InterPro: IPR000805; Glyco_hydro_26.
CC Pfam: PF02156; Glyco_hydro_26; 1.

```

```

DR PRINTS: PR00739; GLHYDRLASE26.
KW Hydrolase; Glycosidase.
SQ SEQUENCE 1021 AA; 115791 MW; A8F7B63109E0F715 CRC64;

```

```

alignment_scores:
Quality: 101.50 Length: 268
Ratio: 0.839 Gaps: 15
Percent Similarity: 45.149 Percent Identity: 24.627

```

alignment_block:

US-09-303-518D-571 x MANA_RHOMR

Align seg 1/1 to: MANA_RHOMR from: 1 to: 1021

```

228 TTTTGGGGAAGCGCAAAATGCGGTTTGGAACTTGCCCGCCGTTTTC 277
33 PheAlaGlyHisGlyLeu.....AlaAspGlnArgLeuIleAl 46
278 AAAACCGGAAGACATCGAAACATGTTCAAGCGGTACAGCGCTGGAA 327
46 AhIstHrGlyArgProAlaLeuAsnAlaGlnArgIleThrGlnGlnLysG 63
328 CAGTGCAGCAGGCTTGGACAAAGCGCAAGGCTGTCTTGCATCAGCGC 377
63 LndArgAspGlnCysArgGlyGlnArgHisArg.....HisHisGln 76
378 GCACATCGCGAGCTACGATTTGG...CGAGCTTACATCAGCGCAGC 424
77 GlyGlyArgAsnLeuArgLysAlaHisArgThrPheHisGlnHisGlnSe 93
425 TTCCGTTTCAACCTGACCGCGCATGTACAAAGCGCGCAAAATCAGATA 474
93 rTnr.....GlnAspGlnAlaHisAspAlaProHisGlyGlnGlnAlaL 108
475 GACAAATCATGACAGC...GGCGAGGCGCGCGGCAAGCAAGCGCC 521
108 ystHrGlyHisGlnGlyLeuGlyHisGlnHisAlaGlnAlaGln..... 122
522 GCCACCGCATACAAAGGCGTCAACAAAT..... 551
123 .....HisGlnGlnGlyGlnSerAsnValAlaAspArgGlnAspGln 136
552 .....CATCAAGCGCGCTGCGCGCGGCGGAGCGCAACC 582
136 yGluProValGlnAlaGlnHisGlnLysAspGlyAlaGlnArgAlaGly 153
583 ATCATCGTCGC.....CGACCAAGCTCCCTTCTCCGCA... 614
153 snAlaProAlaGlyArgValGluLeuGlnGlnGlnProValGlnAlaGln 169
615 .....GGAAGCGCGCGCGCTGTGGG 634
170 HsGlnGlnGlnGlnGlnGlnArgValArgIleGlyLysArgArgGln..... 184
635 CGGATTTTTCGCAAACTGCATACACCATGACACAGCGCGCAAAATGG 684
185 AsnAlaPheAlaProProAlaLeuAsnHisValHisGlyGlyPro..... 199
685 GCACACGTCGAAGCGGTGAACACCTGTCTTTCGTCGCGACGCTGCG 734
200 ..GlyArgLeuGlnArgHisGlyLeuAlaValGlnArgHisValProAla 215
735 .....CGACGCAAGAGCTTCGTTGTCACAA 760
216 ValGlnGlnHisGlnGlnArgValGlnArgGlyArgGlnGlnIleAspH 232
761 T..... 761
232 sValLeuGlyHisGlyLeuProGlyArgGlnArgLeuAlaPheArgAspG 249
762 ..CGCGCCGCTCCAGGGAATTGAACGCAACAAAGCGCCAGATGC... 806

```

1

RP SEQUENCE FROM N.A.
RX MEDLINE=86058903; PubMed=360218;
RA Rothnagel J.A., Weiriel T., Idler W.W., Koop D.R., Steinhert P.M.;
RT "the gene for mouse epidermal filaggrin precursor. Its partial
RT characterization, expression, and sequence of a repeating filaggrin
RT unit.";
RL J. Biol. Chem. 262:15643-15648(1987).
RN (2)
RP REVISIONS.
RA Rothnagel J.A.;
RL Submitted (SEP-1988) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: AGGREGATES KERATIN INTERMEDIATE FILAMENTS AND PROMOTES
CC TERMINAL DIFFERENTIATION AMONG THE INTERMEDIATE FILAMENTS DURING
CC -1- PPM: FILAGRIN IS INITIALLY SYNTHESIZED AS A LARGE, INSOLUBLE,
CC HIGHLY PHOSPHORYLATED PRECURSOR CONTAINING MANY TANDEM COPIES
CC OF 248 AA, WHICH ARE NOT SEPARATED BY "LARGE LINKER". THE
CC PRECURSOR IS DEPOSITED AS KERATOHYALIN GRANULES. DURING TERMINAL
CC DIFFERENTIATION IT IS DEPHOSPHORYLATED & PROTEOLYTICALLY CLEAVED.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: J03458; AAN75559.1; -.
DR PID: A28444; A28444.
DR MGD: MGT:95553; Flg.
KW Phosphorylation; Developmental protein.
FT NON_TER
SQ SEQUENCE 336 AA: 35678 MW: 259F124D3AC0DB2D CRC64.

alignment_scores:	
Quality:	100.50
Ratio:	0.756
Percent Similarity:	46.503
	Length: 286
	Gaps: 10
Percent Identity:	21.329

```
alignment_block:
```

Align seg 1/1 to: FILE_MOUSE from: 1 to: 336

```

36 GCGAACGCCCATGACACATCTCTGTTGTAACGCCCGCTCAATGCCCTCTCC 85
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4 SerGlyArgSerSerAsnArgAspArgProArgInLeuSerProSe 20
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
86 TGCCTGCGCTTCTCTGCTCCACACCGCTGGAAACCGGCTCGACATCTG 136
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
20 rGlnSerSerAspSerGlnValHisSerGly..... 30
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
136 GCGTTTACCTTTAAAGGAAGCCGCCGCCGCAATCGTCGCCAATATCG 186
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
31 ...ValGlnValGlnGlyArg..ArgGlyHis..... 39
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
186 GCAAGCGGGGTTTGAACCCGACAGCAGACGGTTCAAAGCCGTTTGTGGG 236
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
40 ...SerSerSerAlaSerArgArgAlaGlySerSerSerGlySerGlyVa 55
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
236 AAAGCGCAAAATCGGCTTTGCAACTGCCGCCCGCGTTTTCMAAAACCG 286
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
55 LgInGlyAlaSerArgAlaGlyLeuAlaAlaAspAlaSerArgArgSerG 72
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
286 GAAGACATCGAAACAATGTTCMAACGGGTACACGCGCTGGGAACCGTGA 336
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
72 LysAlaArgInGlnGlnAlaSerAlaGlnGlyArgAlaGlySerGlnGly 88
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
336 GCGAGCTTTGGA.....CAAGGCGCAAGGCG 361
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
89 GlnAlaGlnGlnArgValSerSerSerAlaAspArgGlnGlnArgArgGln 105
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

seq_name: SwissProt_40:LIPE_AERHY

seq_documentation_block

ID	LIFE_AERHY	STANDARD;	PRT;	684 AA
AC	P40600:			

DT 01-FEB-1995 (Rel. 31, Created)

DT	16-OCT-2001	Rel. 40	Last sequence update)
DI	01-FEB-1995	(Rel. 31,	Last sequence update)

DE Extracellular lipase precursor (EC 3.1.1.3) (Triacylglycerol 14:0)

GN LIP.
OS Aeromonas hydrophila

OC Bacteria; Proteobacteria; gamma subdivision: Aeromonadaceae

UC Aeromonas.
OX NCBI TaxID=644

RN [1]

RP SEQUENCE FROM N.A.
BC STRAIN-113

Medline=93378423: PubMed=8368830.

RA Anguita J., Rodríguez Aparicio L.B., Naharro G.,
PT "Dunaliella salina" (Hantzsch) Griseb. (Chlorophyta)

expression of an extracellular matrix protein, and RT

human isolate.";

Appl. Environ. Microbiol. 59:2411-2417 (1993)

ESTER HYDROLYSIS AND C6 AND C8 FOR TRIACYLGLYCEROL. HYDROLYSIS

OPTIMAL TEMPERATURE FOR ENZYME ACTIVITY IS 37 DEGREES CELSIUS. HOWEVER AFTER 30 MIN. IT DECREASES.

908 OF THE LIPASE ACTIVITY.

CC -1- CATALYTIC ACTIVITY: triacylglycerol + H(2)O = diacylglycerol + a
CC fatty acid anion.
CC -1- SUBUNIT: MONOMER.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: S65123; AAB28083.2; -
DR InterPro: IPR000734; Lipase.
DR PROSITE: PS00120; LIPASE_SER; 1.
KW Hydrolase; Lipid degradation; Signal.
FT SIGNAL 1 48 POTENTIAL.
FT CHAIN 49 684 EXTRACELLULAR LIPASE.
FT ACT SITE 568 568 CHARGE RELAY SYSTEM (BY SIMILARITY).
SO SEQUENCE 684 AA; 71903 MW; 325418D3936FPA428 CRC64;

alignment_scores: Quality: 100.50 Length: 255
 Ratio: 0.966 Gaps: 14
Percent Similarity: 40.784 Percent Identity: 24.706

alignment_block:

US-09-303-518D-571 x LIPE_AERHY

Align seg 1/1 to: LIPE_AERHY from: 1 to: 684

42 CGCCATGCACATCT...GTGACCGCCCTGCTCAATGCTCTCCT... 86
 ||||| ||||| ||||| ||
272 ArgHisLeuYrProGlyArgAspArgProThrcysLeuProAspProAl 288
87GCTGTGC 93
288 aglyCysArgAlaValProAlaIaGlunIlyAspAlaIleAlaThra 305
94 CTTTCCTGTGTGCACAGCGTGGAAACCGGCTGCGACATCTGCGCTTTA 143
 ||||| ||| ||||| ||
305 laPhe.....GlyProValIleuArgserThrArgLeuLeu 316
144 CTTTAAAGACAGACCGCGCGG..... 167
317 Lys.....ArgProIArgSerIleProValProSerSerCysLeuTh 330
168CATCGTCGCAATATGCGGAGCGGCTTTGAAACCCGAC 207
 ||||| ||| ||||| ||
330 rSerSerProHisArgArgProGlnValProGlyAla...ArgProArgP 346
208 AGCGAGACGGGTAAAGCCGTTTTCGGAAGCGAAAGAAATGCGTTTGA 257
 ||| ||| ||| |||
346 rSerPro.....Gly 349
258 ACTGTCGCCCGCGCTTTTCAAAAACCGAGACATCGAATAATGTCA 307
 ||| ||| ||||| ||
350 ThrValPro.....SerGlnProValArgHisArgGlnCysAlaG1 363
308 AAGCGGTACAGCGTGGAAACAGCTGACAGCGCTTGGACAAAGGCGCA 357
 ||| ||| ||||| ||
363 uGlyValThrArgSerAspArgAlaIleGlyIleGlyValArgGly.... 378
358 GGGCTGCTGTTCATCAGCGCGACATCGGACATGATTTGGGCGGAGC 407
378 378
408 CTACATACGACGACAGCTTCGCTTCACCTGACCGCAATGTACAGCGGC 457
 ||||| ||
379 ...SerGlyProAlaGlyAsp.....Al 385

458 CGAAATATCAAGCATAGACAAATCATGCGAGCGGAGGTGCGCGGC 507
 ||| ||| ||| ||| ||
385 aspCysArgSerAspPropGluArgAlaIaIaGlyArgGly.....G 400
508 AAGCGAAACACCGCCGCCCGCATACAGGGGTCAACAAATCATCA 557
 ||| ||| ||| ||| ||
400 LUGLAlaAspTrpGlyAspAlaHisLeuArgArgGln..... 412
558 GGGCTGCGCGCGCGGAGCGACACCATGCTGCGCGAGCGTCCCT 607
 ||| ||| ||| ||| ||
413AlaIaIaGlyArgAlaGlnHisTrpSerLeuGlnProAlaI 427
608 CTCCGACAGAGAGCGCGCG...CGTGTGGCGGATTTTTCGCAACCT 654
 ||| ||| ||| ||| ||
427 aspAlaIaGlyArgGlyAlaIaIaIaIaIaIaIaIaIaIaIaIaIaI 444
655 GCATACACCATGACACTGCG.....GCCAAAT 683
 ||| ||| ||| ||| ||
444 yspGlnHisHisHisHisHisHisHisHisHisHisHisHisHis 460
684 GGCACACGTCAAGG 698
 ||| ||| ||| ||| ||
461 GlyGlnArgGluArg 465

seq_name: Swissprot_40:FIIL_HUMAN

seq_documentation_block: PRT; 416 AA.
ID FIIL_HUMAN STANDARD; PRT; 416 AA.
AC P20930;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Filaggrin precursor (Fragment).
GN FIG.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_Taxid=9606;
RN 11
RP MEDLINE=89296901; PubMed=2740331;
RA McKinley-Grant L.J., Idler W.W., Bernstein I.A., Parry D.A.D.,
RA Cannizzaro L., Croce C.M., Huebner K., Lessin S.R., Steinert P.M.;
RT "Characterization of a cDNA clone encoding human filaggrin and
RT localization of the gene to chromosome region 1q21.";
RL Proc. Natl. Acad. Sci. U.S.A. 86:4848-4852(1989).
RN 12
RP CITRULLINATION;
RX MEDLINE=96374388; PubMed=8780679;
RA Senshu T., Kan S., Ogawa H., Manabe M., Asaga H.,
RT "Preferential delamination of keratin K1 and filaggrin during the
RT terminal differentiation of human epidermis.";
RL Biochem. Biophys. Res. Commun. 225:712-719(1996).
CC -1- FUNCTION: AGGREGATES KERATIN INTERMEDIATE FILAMENTS AND PROMOTES
CC DISULFIDE-BOND FORMATION AMONG THE INTERMEDIATE FILAMENTS DURING
CC TERMINAL DIFFERENTIATION OF MAMMALIAN EPIDERMIS.
CC -1- PTM: FILAGRIN IS INITIALLY SYNTHESIZED AS A LARGE, INSOLUBLE,
CC HIGHLY PHOSPHORYLATED PRECURSOR CONTAINING MANY "TANDEM COPIES
CC OF 324 AA, WHICH ARE NOT SEPARATED BY "LARGE LINKER". THE
CC PRECURSOR IS DEPOSITED AS KERATOHYALIN GRANULES. DURING TERMINAL
CC DIFFERENTIATION IT IS DEPHOSPHORYLATED & PROTEOLYTICALLY CLEAVED.
CC -1- PTM: Undergoes delamination of some arginine residues
CC (citrullination).
CC
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

EMBL: M24355; AAA52454.1; -

DR PIR: A32947, A32947.
DR MIM: 135940;
DR InterPro: IPR003303; Filaggrin.
DR PRINTS: PR00487; FILAGRIN.
KW Phosphorylation; Citrullination; Developmental protein.
FT NON TER
SQ SEQUENCE 416 AA: 44105 MW: DEEA3218BA043F32 CRC64;

alignment_scores:
Quality: 100.00 Length: 226
Ratio: 0.917 Gaps: 8
Percent Similarity: 48.230 Percent Identity: 20.354

alignment_block:

US-09-303-518d-571 x FILA_HUMAN

Align seg 1/1 to: FILA_HUMAN from: 1 to: 416

```

159 CCGCGCGCGCATCTGCGC.....CAATATCGCGCGCGGTTTGA 199
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
169 ProArgThrSerArgAsnTrpGlySerSerPheSerGlnAspSerAsp 185
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
200 ACCCGACACGACGACGCGTCAACCCGTTTGGGAAAGGCAAAATGC 249
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
185 rGlnGlnHisSerGlnAspSerGlnArgTrpSerGlySerAla..... 199
250 GATTGGAACTTGGCCCGCGCTTTTCAAAAACCGGAGACATCGAAAC 299
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
200 .....SerArgAsnHisHisGly 205
300 AATGTTCAACGCGTACACGCGTGGACACGACGACGCGTTTGAA 349
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
206 SerAlaGlnGlnGlnLeuArgAspGlySerArgHisProArgSerHisG 222
350 AGGCGGAAGGCGTCTGTTTATACGCGCA..... 380
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
222 nGluAspArgAlaGlyHisGlnHisSerAlaAspSerSerArgGlnSerG 239
381 .....CATCGGACGTACGATTGGG.....CGAGCGCTACT 413
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
239 LyrThrArgHisThrGlnThrSerSerGlyGlnAlaAlaSerSerHis 255
414 CAGCCGACGACTTCGTTCCACCTGACCCGACGACGCGCGGAA 463
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
256 GlnGlnAlaArgSerSerAlaGlyAspArgHisGlnSerHisGlnGln 272
464 TCAAGCGATAGACAAATCATGCAAGC.....GGGAGGGTGGC 504
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
272 nSerAlaAspSerSerArgHisSerGlyLeuGlnHisGlnAlaSerS 289
505 GCGAAAGCAAAACCGCGCCACCGG..... 530
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
289 eAlaValAlaArgAspSerGlnHisArgGlyTyrSerGlySerGlnAlaSer 305
531 .....CATCAAGGGGTAAACAAATCATCAAGGCGCTGGCG 568
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
306 AspArgnGlnGlnHisSerGlnAspSerSerPThrGlnSerValSerAlaH 322
569 CCGGCGAGGCAACATCATCTGCGCGACGACGCGCTTCCCGCAGAA 618
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
322 sGlyGlnAlaGlySerHisGlnGlnSerHisGlnGlnSerAlaGlyA 339
619 GCGGCGGCGGTGTC.....GGCGGATTTTTCGCGCAACCTGC 656
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
339 rGSerGlyGlnThrSerGlyHisSerGlySerPheLeuTyrGlnValSer 355
657 ATACACCAT.....GACACTGG 673
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
356 ThrHisGlnGlnSerGlnSerHisGlyTyrThrGlnProSerThrAr 372
674 CCGGAAATTCGACACGCTCAAGGCGT 701
|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

372 gGlyArgGlnGlySerArgHisGlnGln 381

seq_name: SwissProt_40:CCAB_RABIT

seq_documentation_block:

ID CCAB_RABIT STANDARD; PRT; 2339 AA.

AC 005152;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DE Voltage-dependent N-type calcium channel (alpha-1B subunit (Calcium channel, L type, alpha-1 polypeptide isoform 5) (Brain calcium channel II) (BII)).
GN CACNA1B OR CACNA1A5 OR CACNA5.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=93236885; PubMed=8386525;
RA Fujita Y., Mylnieff M., Dirksen R.T., Kim M.-S., Nidome T., Nakai J., Friedrich T., Iwabe N., Miyata T., Furutachi T., Furutama D., Mikoshiba K., Mori Y., Beam K.G.;
RT "Primary structure and functional expression of the omega-conotoxin-sensitive N-type calcium channel from rabbit brain.";
RL Neuron 10:585-598(1993).
CC -1- FUNCTION: VOLTAGE-SENSITIVE CALCIUM CHANNELS (VSCC) MEDATE THE ENTRY OF CALCIUM IONS INTO EXCITABLE CELLS AND ARE ALSO INVOLVED IN A VARIETY OF CALCIUM-DEPENDENT PROCESSES, INCLUDING MUSCLE CONTRACTION, HORMONE OR NEUROTRANSMITTER RELEASE, GENE EXPRESSION, CELL MOTILITY, CELL DIVISION AND CELL DEATH. THE ISOFORM ALPHA-1B GIVES RISE TO THE "HIGH-VOLTAGE ACTIVATED" (HVA) GROUP AND ARE BLOCKED BY OMEGA-CONOTOXIN-GVIA (OMEGA-CTX-GVIA) AND BY OMEGA-AGATOXIN-IIA (OMEGA-AGA-IIIA). THEY ARE HOWEVER INSENSITIVE TO DIHYDROPYRIDINES (DHP), AND OMEGA-AGATOXIN-IVA (OMEGA-AGA-IVA). CALCIUM CHANNELS CONTAINING ALPHA-1B SUBUNIT MAY PLAY A ROLE IN DIRECTED MIGRATION OF IMMATURE NEURONS.
CC -1- SUBUNIT: VOLTAGE-DEPENDENT CALCIUM CHANNELS ARE MULTISUBUNIT COMPLEXES, CONSISTING OF ALPHA-1, ALPHA-2, BETA AND DELTA SUBUNITS IN A 1:1:1:1 RATIO. THE CHANNEL ACTIVITY IS DIRECTED BY THE FORMING AND VOLTAGE-SENSITIVE ALPHA-1 SUBUNIT. IN MANY CASES, THIS SUBUNIT IS SUFFICIENT TO GENERATE VOLTAGE-SENSITIVE CALCIUM CHANNEL ACTIVITY. THE AUXILIARY SUBUNITS BETA AND ALPHA-2/DELTA LINKED BY A DISULFIDE BRIDGE REGULATE THE CHANNEL ACTIVITY.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- TISSUE SPECIFICITY: WIDESPREAD EXPRESSION THROUGHOUT THE BRAIN.
CC -1- DOMAIN: EACH OF THE FOUR INTERNAL REPEATS CONTAINS FIVE HYDROPHOBIC TRANSMEMBRANE SEGMENTS (S1, S2, S3, S5, S6) AND ONE POSITIVELY CHARGED TRANSMEMBRANE SEGMENT (S4). S4 SEGMENTS PROBABLY REPRESENT THE VOLTAGE-SENSOR AND ARE CHARACTERIZED BY A SERIES OF POSITIVELY CHARGED AMINO ACIDS AT EVERY THIRD POSITION.
CC -1- PTM: PHOSPHORYLATED IN VITRO BY CAM-KINASE II, CARX, PKC AND GSK3 (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE CALCIUM CHANNEL ALPHA-1 SUBUNITS FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL: D14157; BAA03202.1;
DR InterPro: IPR002077; Ca_channel;
DR InterPro: IPR002111; Cal_channel_TrpL;
DR InterPro: IPR000636; Cation_chan_non_lig;
DR InterPro: IPR001682; Channel_pore_Ca_Na;


```

712 TTTTCTGCTGGACGCGCTGCGCCGACGACGCTTGGTTGCACAT 761
    :: ||||| ||||| ||| |||
981 .....gluGluproAlaArgHisAlaAlaGHisLysAla 993
762 CCGCCCGTCCAGGGGAAATGACGACACAA 794
    ||||| ||||| |||||
994 ProProlingluThAlaGluLysAspLys 1004

seq_name: SwissProt_40:EVAL_HUMAN

seq_documentation_block:
ID   EVAL_HUMAN          STANDARD;          PRT;          592 AA.
AC   G09502;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Eyes absent homolog 1.
GN   EYAL1.
OS   Homo sapiens (Human)
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX   NCBI_Taxid=9606;
RN   [1]
RP   SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RX   MEDLINE=98030537; PubMed=9361030;
RA   Abdelhak S., Kalatzis V., Hellig R., Compain S., Samson D.,
RA   Vincent C., Levi-Acobas F., Cruaud C., Le Merrier M., Mathieu M.,
RA   Koenig R., Vigneron J., Weissenbach J., Petit C., Weill D.,
RT   "Clustering of mutations responsible for branchio-oto-renal (BOR)
RT   syndrome in the eyes absent homologous region (eyahr) of EYAL1."
RL   Hum. Mol. Genet. 6:2247-2255(1997).
RP   [2]
RP   SEQUENCE FROM N.A. (ISOFORM EYAL1).
RX   MEDLINE=97172972; PubMed=9020840;
RA   Abdelhak S., Kalatzis V., Hellig R., Compain S., Samson D.,
RA   Vincent C., Weill D., Cruaud C., Sably I., Leboyet M.,
RA   Bihner-Glindcz M., Francis M., Lacombe D., Vigneron J.,
RA   Charachon R., Boven K., Bededer P., van Regemorter N.,
RA   Weissenbach J., Petit C.;
RT   "A human homologue of the Drosophila eyes absent gene underlies
RT   branchio-oto-renal (BOR) syndrome and identifies a novel gene
RT   family."
RL   Nat. Genet. 15:157-164(1997).
CC   -1- FUNCTION: MAY BE REQUIRED FOR NORMAL DEVELOPMENT OF BRANCHIAL
CC   ARCHES, EAR AND KIDNEY.
CC   -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; EYAL1 (SHOWN HERE) AND EYAL1B;
CC   ARE PRODUCED BY ALTERNATIVE SPLICING.
CC   -1- TISSUE SPECIFICITY: IN THE EMBRYO, HIGHLY EXPRESSED IN KIDNEY WITH
CC   LOWER LEVELS IN BRAIN, WEAKLY EXPRESSED IN LUNG. IN THE ADULT,
CC   HIGHLY EXPRESSED IN HEART AND SKELETAL MUSCLE. WEAKLY EXPRESSED IN
CC   BRAIN AND LIVER. NO EXPRESSION IN EYE OR KIDNEY.
CC   -1- DISEASE: DEFECTS IN EYAL1 ARE THE CAUSE OF BRANCHIO-OTO-RENAL
CC   (BOR) SYNDROME, AN AUTOSOMAL DOMINANT DISORDER MANIFESTED BY
CC   CYSTS, LACRIMAL DUCT STENOSIS, HEARING LOSS, STRUCTURAL DEFECTS OF
CC   THE OUTER, MIDDLE, OR INNER EAR, AND RENAL DYSPLASIA. ASSOCIATED
CC   DEFECTS INCLUDE ASTHENIC HABITUS, LONG NARROW FACES, CONSTRICTED
CC   PALATE, DEEP OVERBITE, AND MYOPIA. HEARING LOSS MAY BE DUE TO
CC   MONDINI TYPE COCHLEAR DEFECT AND STAPES FIXATION.
CC   -1- SIMILARITY: BELONGS TO THE EYA FAMILY.
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC   or send an email to license@sib-sib.ch).
DR   EMBL: Y10260; CA71309.1; -
DR   EMBL: AJ000097; CA003922.1; -

```

```

DR   EMBL: AJ000098; CA003923.1; -
DR   MIM: 601653; -
DR   MIM: 113650; -
DR   InterPro: IPR001454; Hydrolase.
DR   Pfam: PF00702; Hydrolase; 1.
KW   Developmental protein; Multigene family; Alternative splicing;
KW   Disease mutation.
FT   VARSPLIC 1 41 MEMODLTPSHLSGSSSPGPKGNHSNMPNGTE
FT   VARIANT 487 487 S->P (IN BOR SYNDROME).
FT   VARIANT 505 505 L->R (IN BOR SYNDROME).
FT   SEQUENCE 592 AA; 64593 MW; D62365F81EB652E2 CXC64;

alignment_scores:
Quality: 99.00 Length: 394
Ratio: 0.559 Gaps: 22
Percent Similarity: 44.924 Percent Identity: 20.051

alignment_block:
US-09-303-518D-571 x EVAL_HUMAN ..

Align seg 1/1 to: EVAL_HUMAN from: 1 to: 592

6 TCGTTTACAATTCAGGCTGTTCCCTTGGCAACCGGCATGCACATCC 55
|||||: |||: |||||:
73 SerPheSerProArgProThrHisGlnPheSerProProGlnIleTyrPr 89
56 TGTGACCGCCCGCTCA.....ATGCTCTCCCTGCTGCTGCTTCC 99
|: ||||| |||||: |||||:
89 OsErAsnAlaProTyrProHisIleLeuProThrProSerSerGlnThrM 106
100 TGTCTGCACAGCTGGGAACCGGCTCGACATCGGGCTT..... 140
|||||: |||||: |||||: |||||: |||||:
106 etAlaIaTyr...GlyGlnThrGlnPheThrGlyMeGlnGlnAla 121
141 .....TTACCTTTAAGGAAGACCGCGCGCATCTGTCG 175
||| |||||: |||||: |||||:
122 ThrAlaTyrAlaThrTyrProGlnProGlnProTyrGly...IleS 137
176 CCATATGCGGCA.....GGCGGTTTAAAC 201
|||||: |||||: |||||: |||||: |||||:
137 rSerTyrGlnAlaLeuTyrPalaGlyIleLysThrGlnGlyLeuSerg 154
202 CCGGACAGCAGCGTCAAGCCGTTTGGGGAACGCGCAATG.... 248
|||||: |||||: |||||: |||||: |||||:
154 InsErGlnSerProGlyGlnThrGlyPheLeuSerTyrGlyThrSerPhe 170
249 .....CGTTTGGAACTTGCCCGCGCTTTTCAAAAACCGGAG 289
||| |||||: |||||: |||||: |||||: |||||:
171 SerThrProGlnProGlyGlnAlaProTyrSerTyrGlnMetGlnLys 187
290 ACAT.....CGAAACATGTTCAAGCGGTACAGCGCTGGAAAC 330
|||||: |||||: |||||: |||||: |||||:
187 rSerPheThrThrSerSerGlyIleTyrThrGlyAsnAsnSerLeuTha 204
331 GTGACGAGGC..... 341
204 snSerSerGlyPheAsnSerSerGlnGlnAspTyrProSerTyrProSer 220
342 TTTGGACAAAGGCGCTGCTGTTCATCAGCC.....GC 379
|||||: |||||: |||||: |||||: |||||:
221 PheGlyGlnGlnGlnTyrTyrAlaGlnTyrTyrAsnSerSerProTyrProAl 237
380 ACATCGGACGCTAGATTGGGGGAGCGTACATCAGCGACGCTTCG 429
|||||: |||||: |||||: |||||: |||||:
237 ahIstYrMetThrSerSerAsnThrSerProThrThrProSerThrAsna 254
430 TT.....CCAGCTGACCGGCATGTA...CAACCGCG 458
||| |||||: |||||: |||||: |||||: |||||:
254 IahThrTyrGlnLeuGlnIuProProSerGlyIleThrSerGlnAlaVal 270

```

```

459 GAAATCAAGGATAGACAAATCATGACGGCGGCGGCGGCGGCA 508
      ::::: ::::: ::::: :::::
271 ThrAspProThrAlaGluTyrSerThrIleHisSerProSerThrProI 287
      ::::: ::::: ::::: :::::
509 AAGCGAAACCGCGCCGACCATGACGAGGCGTCAACAAATCATGACG 558
      ::::: ::::: ::::: :::::
287 elYsAspSerAspSerAspArgLeuArgGlySer.....AspG 301
      ::::: ::::: ::::: :::::
559 GCCCTGCGCGCGGCGGAGGC.....AACATCATCTCCGCGACACGCT 602
      ||| ::::: ||| ::::: |||
301 LysSerArgGlyArgGlyArgArgAsnAsnProSerProPro. 317
      ::::: ::::: ::::: :::::
603 CCTCTTCGCGAGAGGCGGCGGCTGGG..... 633
      ||| ||| :::::
317 oAspSerAspLeuArgValArgValPheIleTrpAspLeuAspGluThrIle 334
      :::::
634 .....GCCGATTTTTCGGGCAAA 651
      :::::
334 LeValPheHisSerLeuLeuThrGlySerTyrAlaAsnArgGlyArg 350
      :::::
652 ...CCTGCATACACCATGACACTGCGCGCAAAATGCGACACGCTC.... 693
      ||| :::::
351 AspProThrSerValSerLeuGlyLeuArgMetGluGluMetIlePhe 367
      :::::
694 ....AAGCGGTGAAACCTGTTTCTGTC.....TCGCAAC 727
      :::::
367 eAsnLeuAlaAspThrHisLeuPheHisAspLeuGluGluGlyAspG 384
      :::::
728 GCCGCGCGCGAGGACAGACGCTTGCTGTCATCCGCGCGCGCAAGG 777
      :::::
384 In.....ValHisIleAspAspAlaSerSer 392
      :::::
778 GAATTGAACGCGCAACAAAGCCACGATCCGCGGTTCACCGCATAC 827
      :::::
393 AspAspAsnGly.....GlnAspLeuSerThrTyrAsnGlyTh 406
      :::::
828 CGAA..... 831
      :::::
406 RasPolYpHeProAlaAlaThrSerAlaAsnLeuCysLeuAlaThG 423
      :::::
832 .....TATTGATACGCGGTTTTCGCGACGATAT... 861
      ||| ::::: |||
423 LysAlaArgGlyGlyAlaAspTrpMetArgGlyLeuAlaPheArgTyrArg 439
      :::::
862 .....CTGTTATGTACACCCCTATATA 885
      :::::
440 ArgValLysGluIleTyrAsnThrTyrLys 449
      :::::

```

seq_name: Swissprot_40:CCAB_HUMAN

seq_documentation_block:
ID CCAB_HUMAN STANDARD; PRT; 2339 AA.

```

AC Q00975;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Voltage-dependent N-type calcium channel alpha-1B subunit (Calcium
DE channel, L type, alpha-1 polypeptide isoform 5) (Brain calcium channel
DE I1) (BIII).
GN CACNA1B OR CACNA1A5 OR CACNA5.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OC NCBI_Taxid=9606;
OX [1]
RN SEQUENCE FROM N.A. (ISOFORMS ALPHA-1B-1 AND ALPHA-1B-2).
RP TISSUE=Brain;
RC MEDLINE=92335866; PubMed=1321501;
RX Williams M.E., Brust P.F., Feldman D.H., Patl S., Sherson S.,
RA Maroufi A., Mocce A.F., Veliceljevi G., Ellis S.B., Harpold M.M.;
RT "Structure and functional expression of an omega-conotoxin-sensitive
RT human N-type calcium channel."

```

```

RL Science 257:389-395(1992).
RN [2]
RP SEQUENCE OF 1-94 FROM N.A.
RC TISSUE=Lung fibroblast;
RA Kim D.S., Jung H.H., Park S.H., Chin H.;
RT submitted (Feb-1997) to the EMBL/Genbank/DBJ databases.
CC -1- FUNCTION: VOLTAGE-SENSITIVE CALCIUM CHANNELS (VSCC) MEDATE THE
CC ENTRY OF CALCIUM IONS INTO EXCITABLE CELLS AND ARE ALSO INVOLVED
CC IN A VARIETY OF CALCIUM-DEPENDENT PROCESSES, INCLUDING MUSCLE
CC CONTRACTION, HORMONE OR NEUROTRANSMITTER RELEASE, GENE EXPRESSION,
CC CELL MOTILITY, CELL DIVISION AND CELL DEATH. THE ISOFORM ALPHA-1B
CC GIVES RISE TO THE "HIGH-VOLTAGE ACTIVATED" (HVA) GROUP AND ARE BLOCKED
CC BELONG TO THE "HIGH-VOLTAGE ACTIVATED" (HVA) GROUP AND ARE BLOCKED
CC BY OMEGA-CONOTOXIN-GVIA (OMEGA-CTX-GVIA) AND BY OMEGA-AGATOXIN-
CC ITA (OMEGA-AGA-ITIA). THEY ARE HOWEVER INSENSITIVE TO
CC DIHYDROPYRIDINES (DHP), AND OMEGA-AGATOXIN-IVA (OMEGA-AGA-IVA).
CC CALCIUM CHANNELS CONTAINING ALPHA-1B SUBUNIT MAY PLAY A ROLE IN
CC DIRECTED MIGRATION OF IMMATURE NEURONS.
CC -1- SUBUNIT: VOLTAGE-DEPENDENT CALCIUM CHANNELS ARE MULTISUBUNIT
CC COMPLEXES, CONSISTING OF ALPHA-1, ALPHA-2, BETA AND DELTA SUBUNITS
CC IN A 1:1:1:1 RATIO. THE CHANNEL ACTIVITY IS DIRECTED BY THE PORE-
CC FORMING AND VOLTAGE-SENSITIVE ALPHA-1 SUBUNIT. IN MANY CASES, THIS
CC SUBUNIT IS SUFFICIENT TO GENERATE VOLTAGE-SENSITIVE CALCIUM
CC CHANNEL ACTIVITY. THE AUXILIARY SUBUNITS BETA AND ALPHA-2/DELTA
CC LINKED BY A DISULFIDE BRIDGE REGULATE THE CHANNEL ACTIVITY.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS, ALPHA-1B-1 (SHOWN HERE) AND
CC ALPHA-1B-2; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: ALPHA-1B-1 AND ALPHA-1B-2 ARE EXPRESSED IN THE
CC CENTRAL NERVOUS SYSTEM, BUT NOT IN SKELETAL MUSCLE OR AORTA.
CC -1- DOMAIN: EACH OF THE FOUR INTERNAL REPEATS CONTAINS FIVE
CC HYDROPHOBIC TRANSMEMBRANE SEGMENTS (S1, S2, S3, S5, S6) AND ONE
CC POSITIVELY CHARGED TRANSMEMBRANE SEGMENT (S4). S4 SEGMENTS
CC PROBABLY REPRESENT THE VOLTAGE-SENSOR AND ARE CHARACTERIZED BY A
CC SERIES OF POSITIVELY CHARGED AMINO ACIDS AT EVERY THIRD POSITION.
CC -1- PTM: PHOSPHORYLATED IN VITRO BY CAM-KINASE II, CAPK, PKC AND CGRK
CC (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE CALCIUM CHANNEL ALPHA-1 SUBUNITS
CC FAMILY.
CC -----
CC THIS SWISS-PROT ENTRY IS COPYRIGHT. IT IS PRODUCED THROUGH A COLLABORATION
CC BETWEEN THE SWISS INSTITUTE OF BIOINFORMATICS AND THE EMBL OUTSTATION
CC AT THE EUROPEAN BIOINFORMATICS INSTITUTE. THERE ARE NO RESTRICTIONS ON ITS
CC USE BY NON-PROFIT INSTITUTIONS AS LONG AS ITS CONTENT IS IN NO WAY
CC MODIFIED AND THIS STATEMENT IS NOT REMOVED. USAGE BY AND FOR COMMERCIAL
CC ENTITIES REQUIRES A LICENSE AGREEMENT (SEE http://www.isb-sib.ch/announce/
CC OR SEND AN EMAIL TO license@sib-sib.ch).
CC -----
DR EMBL: M94172; AAA51897.1; -
DR EMBL: M94173; AAA51898.1; -
DR EMBL: U76666; AAC51138.1; -
DR MIM: 601012; -
DR INTERPRO: IPR002077; Ca_channel.
DR INTERPRO: IPR002111; Ca_channel_TrpL.
DR INTERPRO: IPR000636; Cation_chan_non_11g.
DR INTERPRO: IPR001682; Channel_pore_Ca_Na.
DR Pfam: PF00520; Ion_trans; 4.
DR PRINTS: PR00167; CACCHANNEL.
DR IONIC CHANNEL: Transmembrane: Ion transport; Voltage-gated channel;
DR Calcium channel; Glycoprotein; Repeat; Multigene family;
RW Calcium-binding; Phosphorylation; ATP-binding; Alternative splicing.
KW REPEAT 82 359
KW REPEAT 468 712
KW REPEAT 1137 1419
KW REPEAT 1456 1711
FT DOMAIN 1 95
FT TRANSSEM 96 114
FT TRANSSEM 115 132
FT TRANSSEM 133 152
FT TRANSSEM 153 163
FT TRANSSEM 164 183
FT DOMAIN 184 187
FT TRANSSEM 188 206

```


FT	DOMAIN	207	225	CYTOPLASMIC (POTENTIAL).
FT	TRANSSEM	226	245	S5 OF REPEAT I (POTENTIAL).
FT	DOMAIN	246	331	EXTRACELLULAR (POTENTIAL).
FT	TRANSSEM	332	356	S6 OF REPEAT I (POTENTIAL).
FT	DOMAIN	357	482	CYTOPLASMIC (POTENTIAL).
FT	TRANSSEM	483	501	S1 OF REPEAT II (POTENTIAL).
FT	DOMAIN	502	516	EXTRACELLULAR (POTENTIAL).
FT	TRANSSEM	517	536	S2 OF REPEAT II (POTENTIAL).
FT	DOMAIN	537	544	CYTOPLASMIC (POTENTIAL).
FT	TRANSSEM	545	562	S3 OF REPEAT II (POTENTIAL).
FT	DOMAIN	563	573	EXTRACELLULAR (POTENTIAL).
FT	TRANSSEM	574	592	S4 OF REPEAT II (POTENTIAL).
FT	DOMAIN	593	611	CYTOPLASMIC (POTENTIAL).
FT	TRANSSEM	612	631	S5 OF REPEAT II (POTENTIAL).
FT	DOMAIN	632	684	EXTRACELLULAR (POTENTIAL).
FT	TRANSSEM	685	709	S6 OF REPEAT II (POTENTIAL).
FT	DOMAIN	710	1151	CYTOPLASMIC (POTENTIAL).
FT	TRANSSEM	1152	1169	S1 OF REPEAT III (POTENTIAL).
FT	DOMAIN	1170	1185	EXTRACELLULAR (POTENTIAL).
FT	TRANSSEM	1186	1205	S2 OF REPEAT III (POTENTIAL).
FT	DOMAIN	1206	1217	CYTOPLASMIC (POTENTIAL).
FT	TRANSSEM	1218	1236	S3 OF REPEAT III (POTENTIAL).
FT	DOMAIN	1237	1246	EXTRACELLULAR (POTENTIAL).
FT	TRANSSEM	1247	1265	S4 OF REPEAT III (POTENTIAL).
FT	DOMAIN	1266	1284	CYTOPLASMIC (POTENTIAL).
FT	TRANSSEM	1285	1304	S5 OF REPEAT III (POTENTIAL).
FT	DOMAIN	1305	1391	EXTRACELLULAR (POTENTIAL).
FT	TRANSSEM	1392	1416	S6 OF REPEAT III (POTENTIAL).
FT	DOMAIN	1417	1471	CYTOPLASMIC (POTENTIAL).
FT	TRANSSEM	1472	1490	S1 OF REPEAT IV (POTENTIAL).
FT	DOMAIN	1491	1505	EXTRACELLULAR (POTENTIAL).
FT	TRANSSEM	1506	1525	S2 OF REPEAT IV (POTENTIAL).
FT	DOMAIN	1526	1533	CYTOPLASMIC (POTENTIAL).
FT	TRANSSEM	1534	1552	S3 OF REPEAT IV (POTENTIAL).
FT	DOMAIN	1553	1563	EXTRACELLULAR (POTENTIAL).
FT	TRANSSEM	1564	1582	S4 OF REPEAT IV (POTENTIAL).
FT	DOMAIN	1583	1601	CYTOPLASMIC (POTENTIAL).
FT	TRANSSEM	1602	1621	S5 OF REPEAT IV (POTENTIAL).
FT	DOMAIN	1622	1683	EXTRACELLULAR (POTENTIAL).
FT	TRANSSEM	1684	1708	S6 OF REPEAT IV (POTENTIAL).
FT	DOMAIN	1709	2339	CYTOPLASMIC (POTENTIAL).
FT	TRANSSEM	2050	2054	POLY-HIS.
FT	DOMAIN	2118	2122	BINDING TO THE BETA SUBUNIT (BY SIMILARITY).
FT	TRANSSEM	379	396	ATP (POTENTIAL).
FT	DOMAIN	451	458	CALCIUM ION SELECTIVITY AND PERMEABILITY (BY SIMILARITY).
FT	TRANSSEM	314	314	CALCIUM ION SELECTIVITY AND PERMEABILITY (BY SIMILARITY).
FT	DOMAIN	663	663	CALCIUM ION SELECTIVITY AND PERMEABILITY (BY SIMILARITY).
FT	TRANSSEM	1365	1365	CALCIUM ION SELECTIVITY AND PERMEABILITY (BY SIMILARITY).
FT	DOMAIN	1655	1655	CALCIUM ION SELECTIVITY AND PERMEABILITY (BY SIMILARITY).
FT	TRANSSEM	1719	1719	MOD_RES
FT	DOMAIN	1737	1748	CA_BIND
FT	TRANSSEM	256	256	CARBOHYD
FT	DOMAIN	1563	1563	CARBOHYD
FT	TRANSSEM	1675	1675	CARBOHYD
FT	DOMAIN	2164	2339	VARSPIC
FT	SEQUENCE	2339 AA	262494 MW	174A5C6D1E7B39D CRC64.

alignment_scores:
 Quality: 99.00
 Ratio: 0.780
 Percent Similarity: 41.100
 Length: 309
 Gaps: 16
 Percent Identity: 22.330

alignment block:

US-09-303-518d-571 x CCAB_HUMAN

Align seg 1/1 to: CCAB_HUMAN from: 1 to: 2339

```

42  CGCATGACATCCGTTGACCGCCCTGCTCAATGCTCTCCCTGCTG 91
    ||||| ..... ||| ..... |||
803 ArgHisIleuArgPro...AspMetIsthrHisIleuAspArgProLeuVa 818
92  CGCTTCTGCTGTCACACGCTGGGAAACCGCTGGACATCTGGCGTT 141
    : : : : : : : : : : : : : : : : : : : : : : : : : : :
818 IValGluIleuGlyArgAspGlyAlaArg.....GlyProGlu 830
142 TACCTTTAAAGAAAGACCGCGCGCGCATGTCGCAATATGCGGAGCG 191
    : : : : : : : : : : : : : : : : : : : : : : : : : : :
830 aIGlyGlyLysAlaArgProGluAlaGlu.....AlaProGlu 843
192 GGGTTTGACCCCGACGACGACGATCAAAACCGTTTGGCGAAGCG 241
    ||| : : : |||
844 GlyValAspPro..... 847
242 CAATATGCGGTTTGGAATGCCCCCGCTTTTCAAAAACCGGAAGAC 291
    : : : : : : : : : : : : : : : : : : : : : : : : : : :
848 .....ProArg.....ArgHisIsthrGln 854
292 ATCGAACAATGTTCAAGCGGTACACGGCTGGAAACGTCGACGAGC 341
    ||||| : : : : : : : : : : : : : : : : : : : : : : : : :
854 IsArgSplyAspSply.....ThrProIleAlaGly 864
342 TTGGCAAGGCGGAGGCGTGTCTATCAACGCGCA..... 380
    ||| ||| |||
865 ...AspGlnAspArgAlaGluAlaProLysAlaGluSerGlyGluProG 880
381 .....CATCGGACGTAGCATTTGGGCGCA 405
880 YAlaArgGluGluArgProArgProHisArgSer..... 891
406 CGCTACATGACGACGACGCTTCGTTCCACTGACCGCATGTACAAAGC 455
    ||||| : : : : : : : : : : : : : : : : : : : : : : : : :
892 .....HisSerIsgluAlaIleGlyProPro..... 900
456 GCCAAATCAAGCAAGATGACAAATATCATGACGCGGCGAGGTGCGG 505
    ||| : : : : : : : : : : : : : : : : : : : : : : : : :
901 ...GluAlaArgSerGluArgGlyArgGlyProGlyProGlyGlyArg 916
506 GCAAGGCAAAACCGCGCCACCGCATACAAAGGCT.....CAA 546
    : : : : : : : : : : : : : : : : : : : : : : : : : : :
916 GArg.....HisIsthrArgGlySerProGluGlu 927
547 CAATCATCAAGCGCTTCGCGCGGCGGAGCAACATCAT.....CCT 590
    : : : : : : : : : : : : : : : : : : : : : : : : : : :
927 IAlaIleArgGluProArgArgHisArgAlaHisArgHisGlnAspPro 943
591 GCCGACACGACGCTCTCTCCGAGGAAGCGCGCGTGGGCGGATT 640
    : : : : : : : : : : : : : : : : : : : : : : : : : : :
944 SerIysGluSerAlaGlyAlaIleGlyGlyArg..... 955
641 TTTTCGCAAACTGATACACCATGACATGCGCGCAAAATTTGCGACAC 690
    : : : : : : : : : : : : : : : : : : : : : : : : : : :
956 .....AlaArgHisArgGlyGlyProArgAlaGlyPro 967
691 GTCAAGCGGTGAAGAACCTGTTTCTGCTGCGAAGCGCTGCGCGAGG 740
    : : : : : : : : : : : : : : : : : : : : : : : : : : :
967 TgGluAlaGluSerGly.....GluIleProAlaIleArg 978
741 ACAAGCTGCTGTGACATCCGCCCGTCCAAAGGAGATTTGAAGCGCA 790
    ||| ||| |||
979 HisArgIleArgHisIsthrAlaGlnProAlaHisGluAlaValGluIle 995
791 ACAAGCGCACGA..... 803
995 uThrThrGluIsthrGluAlaThrGluIsthrGluIleValGluIleAla 1012

```

804TGCCGCGGTTCACCCGCAATACCGAATATGAT 839
 1012 splyglulysglulubearganshsglnproargluProHsCysASP 1028
 840 ACCGCGTTTCGACGACGATCTGTT 866
 1029 LeuGlutHrSerglYHrValThrVal 1037
 seq_name: SwissProt_40:SPR2_HUMAN
 seq_documentation_block:
 ID SPR2_HUMAN STANDARD; PRT; 221 AA.
 AC 001130;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Splicing factor, arginine/serine-rich 2 (splicing factor SC35) (SC-35)
 DE (Splicing component, 35 kDa) (PR264 protein).
 GN SFRS2
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 67-84.
 RX MEDLINE=92237694; PubMed=1373910;
 RA Fu X.-D., Maniatis T.;
 RT "Isolation of a complementary DNA that encodes the mammalian splicing
 RL factor SC35.";
 RL Science 256:535-538(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Bone marrow;
 RX MEDLINE=92212859; PubMed=1557353;
 RA Vellard M., Sureau A., Soret J., Martinier C., Perbal B.;
 RT "A potential splicing factor is encoded by the opposite strand of the
 RL trans-spliced c-myc exon.";
 RL Proc. Natl. Acad. Sci. U.S.A. 89:2511-2515(1992).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94134745; PubMed=8302870;
 RA Sureau A., Perbal B.;
 RT "Several mRNAs with variable 3' untranslated regions and different
 RL stability encode the human PR264/SC35 splicing factor.";
 RL Proc. Natl. Acad. Sci. U.S.A. 91:932-936(1994).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Lung;
 RA Strausberg R.;
 RL submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP SEQUENCE OF 1-34 FROM N.A.
 RX MEDLINE=93101590; PubMed=1465383;
 RA Sureau A., Soret J., Vellard M., Crochet J., Perbal B.;
 RT "The PR264/c-myc connection: expression of a splicing factor
 RL modulated by a nuclear protooncogene.";
 RL Proc. Natl. Acad. Sci. U.S.A. 89:11683-11687(1992).
 RN [6]
 RP SEQUENCE OF 67-83.
 RX MEDLINE=92249775; PubMed=1577277;
 RA Zahler A.M., Lane W.S., Stolk J.A., Roth M.B.;
 RT "SR proteins: a conserved family of pre-mRNA splicing factors.";
 RL Genes Dev. 6:837-847(1992).
 RN [7]
 RP PROTEIN INTERACTIONS IN SPLICEOSOME ASSEMBLY.
 RX MEDLINE=94084782; PubMed=8261509;
 RA Wu J.Y., Maniatis T.;
 RT "Specific interactions between proteins implicated in splice site
 RL selection and regulated alternative splicing.";
 RL Cell 75:1061-1070(1993).
 RN [8]
 RP BINDING TO U1-70K.

RX MEDLINE=94187841; PubMed=8139654;
 RA Kohltz J.D., Jamison S.F., Will C.L., Zuo P., Luhrmann R.,
 RA Garcia-Blanco M.A., Manley J.L.;
 RT "Protein-protein interactions and 5'-splice-site recognition in
 RL mammalian mRNA precursors.";
 RL Nature 368:119-124(1994).
 RN [9]
 RP RNA BINDING SPECIFICITY.
 RX MEDLINE=95354672; PubMed=7543047;
 RA Tacke R., Manley J.L.;
 RT "The human splicing factors ASF/SF2 and SC35 possess distinct,
 RL functionally significant RNA binding specificities.";
 RL EMBO J. 14:3540-3551(1995).
 RN [10]
 RP SPECIFICITY FOR BETA-GLOBIN MRNA.
 RX MEDLINE=93368668; PubMed=8361546;
 RA Fu X.-D.;
 RT "Specific commitment of different pre-mRNAs to splicing by single SR
 RL proteins.";
 RL Nature 365:82-85(1993).
 CC -1- FUNCTION: NECESSARY FOR THE SPLICING OF PRE-MRNA. IT IS REQUIRED
 CC FOR FORMATION OF THE EARLIEST ATP-DEPENDENT SPLICING COMPLEX AND
 CC INTERACTS WITH SPLICEOSOMAL COMPONENTS BOUND TO BOTH THE 5' AND 3'
 CC SPLICING SITES DURING SPLICEOSOME ASSEMBLY. IT ALSO IS REQUIRED FOR
 CC ATP-DEPENDENT INTERACTIONS OF BOTH U1 AND U2 SNRNPs WITH PRE-
 CC MRNA. INTERACTS WITH OTHER SPLICEOSOMAL COMPONENTS, VIA THE RS
 CC DOMAINS, TO FORM A BRIDGE BETWEEN THE 5' AND 3' SPLICING SITE
 CC BINDING COMPONENTS, U1 SNRP AND U2AF. IN VITRO, BINDS SF2/ASF,
 CC U1-70K AND THE 35 KDA BUT NOT THE 65 KDA SUBUNIT OF U2AF. BINDS TO
 CC PURINE-RICH RNA SEQUENCES, EITHER 5'-AGSAGAGAGA-3' (S-C OR G) OR
 CC 5'-GTTGAGTA-3'. CAN BIND TO BETA-GLOBIN MRNA AND COMMIT IT TO THE
 CC SPLICING PATHWAY.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- ALTERNATIVE PRODUCTS: DIFFERENT FORMS OF THE PROTEIN MAY BE
 CC PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- PTM: EXTENSIVELY PHOSPHORYLATED ON SERINE RESIDUES IN THE RS
 CC DOMAIN.
 CC -1- SIMILARITY: BELONGS TO THE SR FAMILY OF SPLICING FACTORS.
 CC -1- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RRM).
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M90104; AAA60306.1; -
 DR EMBL: X62447; CAA44307.1; -
 DR EMBL: X75755; CAA53383.1; -
 DR EMBL: BC000339; AAH00339.1; -
 DR EMBL: L03693; AAA60162.1; -
 DR PIR: S17328; S17328.
 DR PIR: A42634; A42634.
 DR HSP: P09651; IUP1.
 DR MIM: 600813; -
 DR InterPro: IPR005054; RRM.
 DR Pfam: PF00076; RRM; 1.
 DR SMART: SM00360; RRM; 1.
 DR PROSITE: PSS0102; RRM; 1.
 DR PROSITE: PSS0030; RRM_RNP_1; 1.
 DR Nuclear protein; RNA-binding; mRNA splicing; Alternative splicing;
 KW Phosphorylation.
 FT DOMAIN 14 92 RNA-BINDING (RRM).
 FT DOMAIN 111 116 GLY-RICH (HINGE REGION).
 FT DOMAIN 117 221 ARG/SER-RICH (RS DOMAIN).
 FT CONFLICT 38 38 R -> G (IN REF. 2 AND 3).
 FT SEQUENCE 221 AA; 25575 MW; 9D1B76BDB65701F5 CRC64;

alignment_scores: 98.50 Length: 222
 quality:


```

196 AlaProvalThrSerSerProAlaAlaGlyGlyThrAspGlyAl 212
359 GCGTGGTTCATCAAGCCGACATCCGACATCA.....TTGGGC 402
212 aaIaSerValHisThValSerGlySerProAlaAlaArgLysPheSer 229
403 GAGCGGTACATCAGCAGCAGCTCCGTTCCACCTGACGCGCATGTACA 452
229 hrTherValSerThrArgLysArgThrProProArgTyrArgArg 245
453 GCCGCCGAAAT..... 464
246 LysAlaSerSerProThrThrAlaAlaArgLysArgLysGlyAl 262
465 .....CAAGCATACACAAATCATGACGCGGCGAG.... 497
262 aaSPThAlaThrArgArgSerArgSerThrSerArgLysGlnAlaT 279
498 .....GGTGGCGGCAAGCAAGCAAGCCGCGCCGACCGCATACA 536
279 hrSerArgLysGlyAspArgArgArgArgLysArgSerTyrSer 295
537 AGGGGTAAACAATCATCAAGCCCTGCGCGGCGGCGAGCAACCATCA 586
296 ArgAspThrSerSer.....SerProAspArgLysArgLysArgse 310
587 TCCGCGCCGACGACGCTCTTCCGACGAGAG..... 620
310 rArgGlyGlyProGluThrArgSerGlnSerLeuSerArgSerA 327
621 .....CGCGGCGGTGGCGGATTTTTCGCAACCTGCATACACCAT 665
327 rGserArgSerArgSerArgLysSerSerArgLysAlaAlaPro 343
666 GACACTGCGCGCAAA.....ATTGC..... 686
344 AspaIaValGlyLysSerValArgThrValGlyArgAspHisSerGlyAr 360
687 .....ACAGCTCAAGCGCTGAAACCCCTGTTTCTGCTGC 723
360 gLeuLysArgLeuAspGlnAlaArgAspProProValIleValLeuA 377
724 GAAGCCTGCGCGACGACGACGAG 746
377 rGlyAspaIaAsnLysLeuLys 384

seq_name: SwissProt_40:IE63_HSV2H

```

```

seq_documentation_block:
ID IE63_HSV2H STANDARD; PRT; 512 AA.
AC P28276;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Transcriptional regulator IE63 (VMM63) (ICP27).
GN UL54.
OS Herpes simplex virus (type 2 / strain HG52).
OC Viruses: dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Simplexvirus.
OX NCBI_TaxID=10315;
RN [1]
RP MEDLINE=92113549; PubMed=1662697;
RX McGeoch D.J., Cunningham C., McIntyre G., Dolan A.;
RT "Comparative sequence analysis of the long repeat regions and
RT adjoining parts of the long unique regions in the genomes of herpes
RT simplex viruses types 1 and 2."
RT J. Gen. Virol. 72:3057-3075(1991).
RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: INVOLVED IN THE DOWN-REGULATION OF VIRAL IMMEDIATE-EARLY

```

```

CC GENES AND ACTS IN COMBINATION WITH ICP0 AND ICP4 AS AN ACTIVATOR
CC OF LATE GENES (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO FAMILY THAT GROUPS TOGETHER HSV-1 UL54,
CC HSV-2 UL54, EBV-1 5, VZV 4, EBV BMLF1, HCMV UL69, AND HVS-1 57.
CC -----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D10471; BA01269.1; -.
DR EMBL: Z86099; CAB06702.1; -.
DR PIR: J01498; WMBEXA.
KW Early protein; transcription regulation; DNA-binding.
SQ SEQUENCE 512 AA; 54958 MW; 459651470A503BA7 CRC64;

```

```

alignment_scores:
Quality: 97.00 Length: 135
Ratio: 1.386 Gaps: 9
Percent Similarity: 51.852 Percent Identity: 31.111

```

```

alignment_block:
US-09-303-518D-571 x IE63_HSV2H ..

```

```

Align seg 1/1 to: IE63_HSV2H from: 1 to: 512

```

```

440 CCGCATATCA.....AGCGCGCAAAATC 465
|||||:|||||
98 ProProAlaThrThrGlyAlaThrSerArgLeuGlyThrArgArgSerAl 114
466 AAACGATAGCAAAATCATGACGCGGCGGCGGCGGCAAGGCA 515
114 a.SerProArgLysProHisGlyLysValAlaArgLysGlnProPro 130
516 AACCGCGCCACCGGCGATACAGGGGTCAACAAATCATCAAGCCCTGC 565
131 SerThrLysAlaProHisProArgGly.....GlyArgArgLysArgAr 145
566 GCGCGGCGGAGCAACCATCATCT.....GCCGACACAGCTC 603
145 gArgGlyArgGlyArgGlyArgGlyArgLysAlaAspSerThrProLysP 162
604 CTTCTCCGACGAGAGCGGCGGCGGTGGCGGATTTTTCGCAAGC 653
162 ro.....ArgArgArgValSer.....ArgAsnAla 170
654 TGCAATACATGACACTGCGGCAAAATTCGACACGCTCAAGGCGTGA 703
171 .....HisAsnGlnGlyArgHisProAlaSerAlaArgThrAs 184
704 AAACCGTGTCTTCTGCTGCGACGCGTCCCGGCG.....ACAAAGC 747
184 rGlyPro...GlyAlaThrHisGlyGlnAlaArgArgLysGlyGlnL 200
748 TTCGTTGCGACATCCGCGCGCTCCCAAGGGAATTCAGCGCAAAAGC 797
200 euAspAlaSerGlyLysProArgProArgGlyThrArgGlnAlaProPro 216
798 CCA 800
|||
217 Pro 217

```

```

seq_name: SwissProt_40:YDB4_SCHPO

```

```

seq_documentation_block:
ID YDB4_SCHPO STANDARD; PRT; 297 AA.
AC Q10357;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)

```

DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical 32.2 kDa protein C22E12.04 in chromosome I.
 GN SPAC22E12.04.
 OS Schizosaccharomyces pombe (Fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 CC Schizosaccharomycetales; Schizosaccharomycetaceae;
 CC Schizosaccharomycetes.
 NCBI_TaxID=4896;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=972;
 RA Devlin K., Churche C.M., Barrell B.G., Rajandream M.A., Walsh S.V.;
 RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC DR EMBL; 270043; CAA93891.1; -
 CC DR HSSP; P04355; 4MT2.
 KW Hypothetical protein.
 SO SEQUENCE 297 AA; 32243 MW; 70C128A3BEB2B2BA CRC64;

alignment_scores:

Quality: 96.50 Length: 173
 Ratio: 1.109 Gaps: 10
 Percent Similarity: 50.289 Percent Identity: 23.699

alignment_block:

US-09-303-518D-571 x YDB4_SCHPO ..

Align seg 1/1 to: YDB4_SCHPO from: 1 to: 297

```

109 AGCGTGGGAAACCGGCTCGACATCTGGCGGTTTACCTTTAAG..... 153
    ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
163 SerLeuProAsnTrpIleGlyHisCys.....PheValLeuLysCysVal 177
154 .....GAAGACCGCGCGCATGCTGCCAATATGCGGACGGCGGTT 196
    ||||| ||| ||| ||| ||||| ||||| ||||| ||||| |||||
177 LaspAspSerAspSerAlaTherMetGlyIleIleSerAlaGlySerAlaGlyL 194
197 TGAACCCGACACGACGAGCGTCAAGCGCTTTTGGGAAACGGCAAA 246
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
194 eugly.....GlnAsnThrLysGlnIleCysAlaCysThrGlyLys 207
247 TCCGGTTTGAACCTGCCCCCGCTTTTCAAAAAACCGGAGACATCGA 296
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
208 .SerLeuTrpThrGlnHisAlaGlyLeuLysSerValAsnGlyLys 224
297 AACAAATGTTCAAGCGGTACAGCGCT.....GGGACACGCGTC 334
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
224 eCysCysSerLysAspSerSerProSerGlyLysProSerGlyCys 240
335 AGCAGGCTTTGGACAAGGCGGAGGCTGCTGTTTCATCAGCGGACATC 384
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
241 SerGln.....GlyLysSerCysSerLysLys..... 252
385 GGCACCTACGATTTGGCGGACGCTACATCAAGCCAGAGCTTCGTTCCA 434
252 ..... 252
435 CCTGACCCCATGTACAGCCGCGCAAAATCAAGCGATAGACAATAATCA 484
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
253 .....ProSerCysSerGlnGlyLys.....LysGly 262
485 TGC.....AGGCGGACAGGAGTCCGCGCAAGGCAAAACCGCGCC 525
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
263 CysCysSerThrGlyLysThrSerCysSerGlnGlyLys..... 276

```

526 ACCGGCATCAAGGGGCTCAACAAATCATCAAGCGCTCGCGGGCGCA 575
 277
 576 GGCAACCATCATCTCGC 592
 283 LuLysProSerCysCys 288

seq_name: SwissProt_40:CRE1_NEUCR

seq_documentation_block:

ID CRE1_NEUCR STANDARD; PRT; 430 AA.

AC 059958;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DE DNA-binding protein cre-1 (Carbon catabolite repressor).

GN CRE-1.

OS Neurospora crassa.

OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;

OC Sordariatales; Sordariaceae; Neurospora.

NCBI_TaxID=5141;

RN [1]
 RP SEQUENCE FROM N.A.

RC STRAIN=74-OR23-1A;

RA de la Serna I., Tyler B.M.;

RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.

CC -I- FUNCTION: INVOLVED IN CARBON CATABOLITE REPRESSION. REPRESSIONS THE

CC TRANSCRIPTION OF A NUMBER OF GENES BY BINDING TO A GC-RICH REGION

CC IN THEIR PROMOTER (BY SIMILARITY).

CC -I- SUBCELLULAR LOCATION: Nuclear (By similarity).

CC -I- SIMILARITY: BELONGS TO THE CRE1/MIG GROUP OF C2H2-TYPE ZINC-

CC FINGERS PROTEINS.

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

```

DR EMBL; AF05464; AAC13555.1; -
DR HSSP; P07248; IABD.
DR InterPro: IPR000822; Znf-C2H2.
DR Pfam: PF00066; Zf-C2H2; 2.
DR PRINTS; PR00046; ZNCFINGER.
DR SMART; SM00355; Znf_C2H2; 2.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 2.
DR PROSITE; PS0157; ZINC_FINGER_C2H2_2; 2.
KW DNA-binding; Transcription regulation; Repressor; Zinc-finger;
KW Nuclear protein; Metal-binding;
FT ZN_FING 78 100 C2H2-TYPE.
FT ZN_FING 106 130 C2H2-TYPE.
FT DOMAIN 36 45 POLY-GLN.
FT DOMAIN 142 146 POLY-GLN.
SO SEQUENCE 430 AA; 46954 MW; FDD304889A7D8085 CRC64;

```

alignment_scores:

Quality: 96.00 Length: 297
 Ratio: 0.857 Gaps: 12
 Percent Similarity: 37.710 Percent Identity: 22.222

alignment_block:

US-09-303-518D-571 x CRE1_NEUCR ..

Align seg 1/1 to: CRE1_NEUCR from: 1 to: 430

```

8 GTTACATTCAGGCGTGTTCCTTTTGGCAACGCCCA..... 46
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
150 ValHisAsnHisGlyLeuGlnProAspMetMetProProGlyProly 166

```

```

47  TGACATCTCTGTACCGCCCTGCTCAATGCTCTCTGCTGCT 95
166  salalaleargseralaprothralametserprobsanvalserp 183
96  TTCCGTCTGTCACAGCTGGGAACCGGCTCGACATCTGCGTTTAC 145
183  ropro.hisserlyserprotyralsphenelaproserglyleuanhr 199
146  TTTTAAAGAAAGACCGCGCCCATCTGCCCAATATCGCCAGCGGCT 195
199  cyr.....serhiserargseralalyser... 209
196  TTGAACCCCGACAGCAGCGTCAAAACCGTTTTCGGAAACCGCAA 245
209  ..... 209
246  ATGCGGTTTGAACCTGCCCGCGTTTTCAAAAACGGAGACATCG 295
209  ..... 209
296  AAACATGTTCAAACCGCTACACGCTGGGAACATGACAGGCTTG 345
210  .....Glnserglyproaspleserleuallargalalaag 223
346  GACAAAGGCGAAG.....GCTGCTTCATCAGCGCGACATCGCGAG 389
223  Tglnvalglurargaspalyalalalalshishphedlnproarghe 239
390  CTACGATTTGGCGGAGCTACATCAGCCAGCAGCTTC.....GT 430
240  Glnphetrglyasnthreuhisalaalathralaserarganghle 256
431  TCCA.....CGTGAACCGCATGTCAACGCGCGCAAAATCAACGATA 474
256  uproglyleuglnalalythhiserlyserhiserhishgluaph 273
475  GACAAATCATGACAGCGGAGGCTGCGGCAAAAGCAACCGCGCC 524
273  lsasprshistrgly.....Glnserlytrghisala 284
525  CACCGGCATACAAAGGGGTCAAAATCATCAAGGCCCGCGCGGGGG 574
285  Lysargserargprobsanserprobsanserthralaproserproth 301
575  AGGCAACATCATCTCGCCGACAGCTCTTCGCGAGAGAGCGGC 624
301  rpheserhishasprserleuserprothpro..... 311
625  GCGGTGGGCGGATTTTTCGGAACCTGCATACACCATGACACTGCG 674
311  ..... 311
675  GGCAGAAATTGGCACAGCTCAAGGCGTGAACCCGTGTTTCTGCTGC 724
312  .....Asphisthrpro.....LeuAl 317
725  AACGCTGCGCGACAGCAAGGCTGCTGTCACATCG..... 764
317  athrproalalalalalalalalalalalalalalalalalal 333
765  .CCCCGTCCAAAGGAATTGAACGCAACAAACCCAGATGCGCGCT 812
333  eurpropropharganleuserleuglylndlnhishthtrproala 349
813  GTTCAACCGCAATACCAATATGATACCGCTTTC 851
350  leuthrproleuglnulproalaleuasprglylndlnhishthtr 362
seq_name: SwissProt_40:NONA_DROME
seq_documentation_block:
ID NONA_DROME STANDARD; PRT; 700 AA.

```

```

AC 004047;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE No-on-Transient A protein.
GN NONA.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_Taxid=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90262721; PubMed=2344408;
RA Jones K.R., Rubin G.M.;
RT "Molecular analysis of no-on-Transient A, a gene required for normal
RT vision in Drosophila."
RL Neuron 4:711-723(1990).
CC -!- FUNCTION: REQUIRED FOR NORMAL VISION IN DROSOPHILA.
CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; I (SHOWN HERE) AND II; ARE
CC PRODUCED BY ALTERNATIVE SPLICING.
CC -!- SIMILARITY: CONTAINS 2 RNA RECOGNITION MOTIFS (RRM).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL: M33496; AAA03214.1; -.
DR EMBL: M33496; AAA03215.1; -.
DR HSSP: P19339; ISXL.
DR FLYBASE: FBgn0004227; nona.
DR InterPro: IPR000504; RRM.
DR Pfam: PF00076; rrm; 2.
DR SMART: SM00360; RRM; 2.
DR PROSITE: PS50102; RRM; 2.
DR PROSITE: PS00030; RRM_RNP_1;
KW RNA-binding; Repeat; Vision;
FT DOMAIN 302 374 RNA-BINDING (RRM) 1.
FT DOMAIN 376 457 RNA-BINDING (RRM) 2.
FT DOMAIN 73 76 POLY-GLY.
FT DOMAIN 254 264 POLY-GLY.
FT DOMAIN 619 626 POLY-GLY.
FT DOMAIN 648 652 POLY-ASN.
FT VARSPLIC 666 700 DSFAFEFGVNNMNOGNGNGNGGNNVPGRRRF -> VC
FT PPKRYPTKYSVTNRSEVLEIQLSMIPMKL (IN
FT ISOFORM II).
SQ SEQUENCE 700 AA; 76967 MW; 732DB77FC5DFED47 CRC64;

alignment_scores:
Quality: 95.50 Length: 221
Ratio: 0.910 Gaps: 10
Percent Similarity: 47.511 Percent Identity: 23.529

alignment_block:
US-09-303-518D-571 x NONA_DROME ..

Align seg 1/1 to: NONA_DROME from: 1 to: 700

183 GCGGAGCGGGGTTGAAACCGCACAGCAGCGTCAAGCGGTTTTC 232
|||||
77 Alavalglygly.....Prosnlnshlnysasnph 87
|||||
223 CGGAACCGCAAAATGCGGTTTGAACCTGCCCGCGGTTTTCAAAAA 282
|||||
87 eglYshnshlnysglYlph.....ValG 96
|||||
283 CGGAGACATCGAACAATGTTCAAGCGGTACAGCGTGGGACAC... 329
|||||

```

```

96  lYAsnArGAsnArGAsnAsnArgAlaGlyAsnGlnAsnArgThrPhe 112
330 .....CGTGCAGCAGGCTTGG 346
113  ProGlyAsnAsnSerAsnGlnLysProAsnAsnGlnThrSerLysAl 129
347  ACAGGGCGCAGAGGCTGCTTCATCAGCGCGCACATCGCAGCTACGAT 396
129  aAspGlyProAsnAlaLeuAlaLysAsn..... 138
397  TTGGGGGAGCGCTACATCAGCCAGCAGCTCCGTTCCACTGACCGCCAT 446
139 .....AsnGlnProAlaThrAlaAla..... 145
447  GTACAGCCGCGCAAAATCAAGCGATAGACAATCATGACGCGGGCA 496
146 .....AlaGlyGlnAsnGlnAlaAsn...GlnAsnAlaAsnLysGly 159
497  GGGTGGCGGCAAGGCAAAACCGCCGCGCATACAGGGGTCAAA 546
159  nAsnGlnArgGlnGlnAsnGlnAsnGlnAsnGlnValHisGlyGlnG 176
547  CAATCATCAAGGCCCTGCGCGC.....GGG 572
176  lYAsnGlnGlyLysProGlyAsnGlnGlyAlaGlyAsnGlnGlyGly 192
573  CGAGGCAACCATCATCTGCGCCGACGATCCCTTCGCCGAGAGAGCG 622
193  GlnGlyAsnGlnGly.....GlyAlaGlyAsnGln 202
623  GCGCGCTGTGGCGGATTTTTCGGCAAACTGCATACACCATGACACTG 672
202  nGlyAsnGlyGlnGlyPheArgGlyArgAsnAlaGlyAsnAsnGlnGly 219
673  GCGCAAAATTGGC.....ACAGTCAAAGCGGTGAAAACCTGTTT 716
219  LysGlyPheSerGlyGlyProGlnAsnGlnGlnArgAspAsn..... 232
717  CTGCTGGAGACGCGCTGCGCAGCAGACAGGCTTCGTGTCACATCCGCC 766
233 .....ArgAsnArgSerGlyPr 238
767  CCGTCCAAGGGGA 779
238  oArgProGlyGly 242

```

